**Higher proportion of positively-linked language in female university students reviewing the competency of scientific educators**

**For:**

**Lisa Formentini**

**Farah Diba Khan**

**Imran Il Gouncci**

**Marine Menardin**

**Carlos Liévano**

**Jianqiu Kang**

## **Abstract**

Gender politics are a hotly debated topic among the social sciences. With a large number of studies often presenting contradictory findings on identical topics, it seems that consensus among experts has not yet been established. Here, we discuss the possibility of gendered differences in Parisian university students’ responses when reviewing the competency of a scientific educator and the implications of our results. However, of the many metrics we analyzed, it was established that the small sample size made drawing of significant conclusions difficult apart from a single metric; The choice of words. Females in this study were shown to tend more toward positive-linked words than their male counterparts, which may have some implications for understanding the differences in perception of male and female students of differentially gendered scientific educators.

## **Introductio**n

Gender politics has played an increasingly prominent role in public discourse worldwide. With new initiatives and organizations focused on investigating, raising awareness of, and combating the effect of gender bias springing up and challenging the social norms in recent times, this issue will likely only increase in relevance over the coming years (Huang J. et al. 2020). The fields of Science, Technology, Engineering, and Mathematics (STEM) are areas where systems for determining the existence of negative gender stereotypes could potentially lead to a better work environment (Osborne JW. 2001).

The term ‘gender role’ refers to the appropriate position of a specific gendered individual in a certain society, which is influenced by social expectations and regulated by societal norms. Many studies underline the existence of gender-related differences in teaching (L. Wan, 2019; X. Li, 2018) for example, in China, where 34% of the middle school teachers are women, they are described as more likely to find creative ways to teach, distinct from ordinary procedures, which are more frequently used by male teachers ( L. Yao, 2018). Moreover, Hofstede (1986) underlined that cultures have different points of view on teachers, depending on their belonging to an individual or collective culture. This study implicates the need to choose a specific culture in order to obtain valid results when studying gender bias.

The existence of gender roles in science is a long treated subject in cognitive and social psychology. Several studies focused on the perception of women in STEM have suggested that women are on average seen as less competent in scientific fields (Geary & Stoet, 2018; Leyens, 1997, Jones Howe and Rua 2000). There is some evidence in the literature that indicates that children are primed for pursuing social scientific fields while men are primed towards pursuing STEM (Pla-Julian & Diez 2018). A well-known experiment pertaining to observational learning is the 1963’s Bobo Doll experiment, by Bandura, Ross, and Ross. It highlighted that social behavior is able to be transferred to children, and reinforced at a young age, largely independent of the nature of the behavior. There is also some evidence that behavior that is taught in childhood persists throughout a child's development and is expressed in adulthood. (Kaiser FG et al. 2018)

In this study, we attempt to shed light on the potential presence of gender stereotypes in Parisian university students with European cultural backgrounds. We aim to do this by investigating their perception of the competence of science educators based on their gender. Our guess is that female scientific educators will be perceived as less competent than males when rated by Parisian university students with European cultural backgrounds. Gender stereotypes are not typically socially acceptable in occidental societies (Pentland A. et al 2013). Therefore, we will assume that only a vast minority of participants will readily or intentionally display stereotypes and gender perception differences. Following this, the need for a standardized method of bias-detection is apparent. In this study, the brief social desirability scale (BSDS) was utilized to provide information regarding the participant's potential for bias. We chose the BSDS, because it is very brief, and it has great validation scores.

## **Methods**

In the methodology to try to measure the perception of Parisian university students with European cultural backgrounds about the competence of science educators based on their gender, we used two tools with the participants. The first, an audio podcast, we consider that by listening to a speaker about a technical scientific topic, people can have a perception of their competition, without falling into biases by image, stereotype, age and others. The second, and after listening to the audio-podcast, a questionnaire was applied with different questions that will yield data according to our objective and hypothesis, for further analysis.

### **Research tool, the audio-podcast**

In order to investigate the possible gender’s bias of perception, the participants will be shown two near-identical self-made audio podcasts about the same topic: “Science of the Gaps”. In order to not distract the participants and to limit the bias, we decided to use only the audio of our recordings.

Concerning the speakers in those audio (one per audio), they will read the same text, explaining the same notions, with the same sentences, using the same words and having the same english level and pronunciation. The only difference is that initially the participants will choose a group, group A with a male speaker and group B with a female speaker, but they will all access the same questions.

Concerning the educators talking, they are part of our research team: they both have scientific background and they built the audios together.

Concerning our subject, it has been chosen because it allows us to be slightly more technical than what is generally considered layman knowledge. Moreover, displaying technical audios can aid in ensuring all participants are equally new to the topics being presented, which can reduce the bias of our results. To reduce bias even further, the audios will have to have an equal tone, accent, progression and presentation. The idea is that the sole influencing factor in the participants' opinions are to be the gender of the presenter, and nothing else.

### **Research tool, questionnaire.**

After the participants will have listened to the audio, they will be asked to complete an online questionnaire for both groups (Group A & B) and with the following elements.

Questionnaire available in this link: [**https://forms.gle/uZqDPSLXGRHgHGHQA**](https://forms.gle/uZqDPSLXGRHgHGHQA)

1. Introduction: Presentation of the topic with a short explanation. Despite being a technical topic, presenting it more colloquially. Here the participant chooses the group in which she is going to participate, group A or B.
2. Questions to evaluate the comprehension of the audio: With the idea of determining whether people heard the audio and understood its content, three evaluative questions were asked about the scientific topic of the podcast.
3. Questions about interest for the topic: The objective of these questions was to determine if the person had prior knowledge of the subject, if he was interested in listening to it and if the speaker was able to influence that possible interest.
4. Questions of perception and competence of the speaker: These questions are central to our research, their objective is to directly and indirectly measure the perception and competence of the speaker and it is where we could more closely determine some type of gender bias. This, through a scale that allows a quantitative analysis, but also an open question that yields valuable qualitative data.
5. How would you categorize the speaker’s competence in the topic discussed on the audio? “Below is a linear scale ranging from 1 (not competent) to 7 (extremely competent).
6. You are the founder of a tech start-up in Silicon Valley. You have advertised 2 open positions in your physics team on LinkedIn. In the advertisement you state that the individuals interested in applying for a job should email a self-produced, educational, and entertaining physics video. If you had received an email with the audio you heard previously, how likely is it for you to invite the speaker to an interview? “Below is a linear scale ranging from 1 (not likely) to 7 (highly likely).
7. Open question: “The person you saw presenting is going to an interview into the university you're studying at to become professor. If you know you're taking their course next semester, would you like them to be hired and why ?” (open answer).
8. Questions for the social desirability scale: The questionnaire will contain a social desirability scale, in order to eliminate the social desirability bias. In fact, this psychology scale is very used in order to prevent the participants from lying in their answers, and to prevent our results from being compromised. We will be using the BSDS (Brief Social Desirability Scale), which is inspired from the Marlowe-Crowne inventory, but it is reduced to 4 or 5 items. In fact, Marlowe-Crowne is described as “ relatively lengthy and as a result leads to lack of participants’ co-operation” (Haghighat, 2013), and we wanted to avoid this.

f. Socio Demographic questions: The sociodemographic questions allow crossing the main variable on gender bias with important elements such as sex, educational level, academic backgraund and of course, the cultural identity of the participants.

1. Age, open question.
2. Sex, between “Female, Male”
3. The current level of education.
4. Academic background.
5. Cultural identity

### **Data gathering and processing**

1. Participants and application

The target population of our research was centered on Parisian university students with a European cultural background. Therefore, the search for participants was focused on this population, although it limited the number of participants, it also kept the objective and result of the research delimited.

The call was sent by electronic media with the link of the survey, clarifying that it consisted of listening to a 2-minute podcast and carrying out a survey of maximum 10 minutes, there the participants first had to choose a random group, A or B, and according to this The choice was directed to listen to podcast A with the male speaker or b with the female speaker. After listening to the podcast, the survey directed them to answer the same questions, regardless of the group.

1. Data gathering tool

The main means of dissemination of the survey was the institutional mail of the Center de Recherches Interdisciplinaires (CRI in short), but also, the members of the group use facebook groups and direct contact with university students.

The information collection tool was google forms, which allows building surveys to be sent by a link to fill out from a computer or smartphone. All the answers to the questionnaire were saved in this application which allows them to be later exported in different formats such as Excel, for later analysis.

1. Data access, storage and processing:

The datas of the online survey was stored as an excel sheet, directly downloaded from the Google form (see annexes). In the processing stage, the data of the results was organized, aggregated, and analyzed. This is the stage where we make assumptions about our data pre-validation. Then, after we selected the participants, we analyzed the validity of our data with a t-test.

### **Methodology for the data analysis**

1. **Data filtering:**

In this section, we will filter out the participants that don’t fulfill our criterias, which include their learning institute, cultural background, social desirability, content understanding and science background.

First, we will make sure that all participants are students in parisian institute. Each participant has to answer “Yes” at the following question: “Are you currently a student in a parisian institute?”. We will have to remove them from our further analysis if they answer “no”.

Next, we will be searching for an european background. If a participant does not have that background, we will have to cut his data out of our statistical analysis. The correlated question to this parameter is “ What culture do you identify with the most?”, and the accepted answers are “north european, east european, west & central european, south european”.

Then, we use the BSDS scale to test if the participant is answering honestly to the survey. Only participants that score at least 2 out of 4 on the BSDS scale were kept for further analysis. The first 3 questions “Would you smile at people every time you meet them?”, “Do you always do what you preach?”, and “Do you always keep your promises no matter how inconvenient they might be?” need to be answered “no”. The last question “Would you ever lie to someone?” needs to be answered “yes”. If a participant is scoring more than 2/4 at those questions, he should be evicted.

Thirdly, our participants will have to obtain at least 2 correct answers on the 3 content questions: “How is torque created from bending spaghetti?”, “What are the Kirchhoff rod equations?” , “What is the cause of the flexural wave described by the video?” .

Lastly, we will check if our participants all come from a scientific background. The linked question in our questionnaire is “What is your academic background?”, and the expected answers are “natural science, social science, technology, psychology, engineering, mathematics” or “other” with a scientific answer.

1. **Main data analysis:**
2. **Multiple regression**

We are trying to investigate a connection between the speaker’s gender and its perceived competence level. This could help highlight actual gender inequalities in the world of science, and help inform the populations.

On one side, we will do a quantitative statistical analysis: a multiple regression.The regression is a test describing the relationship between dependent and independent variables. It will provide us a R^2 for each dependent variable: this indice represents the percentage of the variance of the dependent variable explained by each independent variable. The coding principles are shown in the following table.

First, we will be measuring a list of categorical independent variables. The factor we will be investigating for our hypothesis is the gender of the speaker, the age of the participants, the current level of education of the participant, the cultural background, the familiarity of the participant and the interest of the participant in the topic.

Secondly, we will be measuring different dependent numerical variables linked to the perceived competence, including both abstract perceived competence and actual judgment of competence. Abstract perceived competence is naming the “competence” word and not contextualizing. While, actual judgment of competence is not naming the “competence” word, but contextualizing and involving the participants.

Table 1. Multiple regression coding principles

|  |  |  |  |
| --- | --- | --- | --- |
| Factors types | Factors names | Answers | Codes |
| Categorical independent variable | the gender of the speaker | Group A (male) | 1 |
| Group B (female) | 2 |
| the age of the participants | age | number |
| the current level of education of the participant | Bachelor/Licence | 1 |
| Master | 2 |
| PhD/Doctorat | 3 |
| Post-Doc | 4 |
| the cultural background | North European | 1 |
| East European | 2 |
| West & Central European | 3 |
| South European | 4 |
| East Asian | 5 |
| West Asian | 6 |
| South Asian | 7 |
| Middle Eastern | 8 |
| South American | 9 |
| North American | 10 |
| West African | 11 |
| East African | 12 |
| South African | 13 |
| North African | 14 |
| other | 15 |
| Binary categorical independent variable | the familiarity of the participant | Yes | 1 |
| No | 2 |
| the interest of the participant in the topic | Yes | 1 |
| No | 2 |
| Dependent numerical variables | Abstract perceived competence | not competent-extremely competent) | 1-7 |
| Actual judgment of competence | not likely-highly likely | 1-7 |

We will run a multiple regression test per dependent variables, linked to each independent factor that we previously mentioned. This will help us to identify the impact of the gender of the speaker. In order to do those calculations, we will use a free and open-source software called JASP, used to do statistical analysis. This software will provide multiple regression and descriptive analysis, as well as illustrations and plots of our datas. Our survey will be done on google form, which will provide us the data as an excel sheet or a CSV sheet. This CSV sheet is the one we will have to use on the JASP software.

Based on prior studies, we expect that having we will provide the following results: The ”woman speaker” group and the “man speaker” group will be rated differently, and the group “woman” has lower rating in terms of perceived competences then the “men” group. This should be happening in both actual competence judgment and abstract perceived competence situations. We do not have any hypothesis regarding the other measured factors.

1. **Correlations**

In addition to the multiple factors impacting the judgment of one’s competency and the importance of gender in this judgment, we will observe some other possible effects. This could help highlight possible explanations of our results that we did not see at first sight, and it could underline other exciting effects other than the gender.

We will do another statistical quantitative analysis: correlations. The correlation is a statistical test, usually visualized with a heat map, describing the link between two different factors, without showing any causality (while the regression shows causality from one variable on another). The variables will be studied by pairs, which are shown in table 2.

We will use a free and open-source software called JASP to do statistical analysis. This software will provide correlation analysis. Our survey will be done on google form, which will provide us the data as an excel sheet or a CSV sheet. This CSV sheet is the one we will have to use on the JASP software. In order to present a visualization of those results, we will use Jupyter Notebook and create a heatmap. It will present every correlation between every variable by a degree of color: the darker, the closer to 1 the correlation is, the clearer, the closer to -1 the correlation is.

Table 2. Correlations coding principles

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Variables pairs | Questions | Variables types | Answers | Codes |
| The link between the initial interest and the final interest | Does this topic interest you? (Initial interest) | Binary categorical variable | Yes | 1 |
| No | 2 |
| To what degree has the speaker interested you in the topic discussed in the audio? (Final interest) | numerical variable | Not at all-Very much | 1-7 |
| The link between the rating of the speech as a student or a teacher | Student judgment | categorical variable | A, B, C, D, or E | Alphabet |
| Teacher judgment | categorical variable | Not at all-Very much | 1-7 |
| The link between the rating of the speech as a student or a professional employee | Student judgment | categorical variable | A, B, C, D, or E | Alphabet |
| Professional employee judgment | categorical variable | Not at all-Very much | 1-7 |
| The link between the rating of the speech as a teacher or a professional employee | Teacher judgment | categorical variable | Not at all-Very much | 1-7 |
| Professional employee judgment | categorical variable | Not at all-Very much | 1-7 |
| The link between the initial familiarity and the understanding | How much of the content do you feel you understood? (understanding) | categorical variable | Not at all-Very much | 1-7 |
| Were you familiar with this topic before hearing the podcast? (Initial familiarity) | binary categorical variable | Yes | 1 |
| No | 2 |
| The link between the final interest and the understanding | To what degree has the speaker interested you in the topic discussed in the audio? (Final interest) | numerical variable | Not at all-Very much | 1-7 |
| How much of the content do you feel you understood? (understanding) | categorical variable | Not at all-Very much | 1-7 |

1. **Thematic analysis**

On the other side, we will do a qualitative analysis via a thematic analysis of content on the Open Question : “The person you heard presenting is going to an interview into the university you're studying at to become professor. If you know you're taking their course next semester, would you like them to be hired and why ?”. This analysis will be very useful for us in order to investigate other possible arguments or bias in our study, forcing participants to answer in a way or another. We will use those informations in our conclusions, our critics, and our results analysis. Those answers could even explain our main results if they underline some major bias.

We will split the participants into 2 groups, either male or female speaker, and in 3 subgroups, depending on their answer: the “yes i would hire the speaker” group, “no I won’t” and “I don’t know”.

We will take a closer look at every participant’s answer in order to have a first look at their critism. Next, we will calculate the frequency of appearance of each word used by the participants, next to having put aside the stopwords. Stopwords “english” is a python function repertoring all the most common words in the english language that do not have any meaning by themselves, for example: “I”, “have”, “‘ll”, etc. This list can be found freely on the internet. Afterwards, we will extract the stem of the most used words in the participant’s answers in order to find the most recurring themes being underlined in this open question. A stem is the core of a word on which prefix and postfix can be added, for example chang is the stem of changing and changer. The associated python function is Stemmer and can be found freely online. We will next compare the stems presented in each subgroup.

For the purpose of the present analysis, there exists no definitive theoretical justification for ordering the entry of critical item variables and list variables in either of our statistical analysis, so we opted for the simultaneous entry of all variables in a single step for the regression and the correlation, and a random order for the content analysis.

## **Results**

In order to proceed to the statistical analysis, we extracted the answers of the 32 participants (53,125% of male participants) to our questionnaire. We included all our participants in the study. We recoded every column that had a textual answer to pick into a numerical column. The table 1 shows how the different questions have been renamed, and what each numeric answer means. The breakdown and descriptive analysis on each question are in the table 2.

**Table 1**

https://docs.google.com/spreadsheets/d/1d6ztp2PT3pMwn3IF\_TtGnoaYW201yEqGIGQlMPB\_euk/edit?usp=sharing

**Table 2:** https://docs.google.com/spreadsheets/d/1G7GIAkkQxH1bPPGLt51mAh0KxYS-KicK6MJlTLzOvi4/edit?usp=sharing

We kept the participants with missing datas, and we filled them as much as possible, ex: participant 1 forgot to fill the Gender of the speaker question. When it was not possible, we left an empty answer, ex: missing answers of the Open question.

We consider first the multiple regression and its findings, then we turn to the correlation analysis between our 11 factors, and we end with a content analysis.

In the multiple regression analysis, we analysed the answers to 11 questions, called “covariates”, in opposition to the answers to the 4 speaker’s competency evaluation questions, called “variables”. Each variable has its own regression analysis, and for each regression analysis, all the covariates were entered simultaneously.

The covariates were the Gender of the speaker (R), the Feeling about the understanding, the Audio as coming from a student (R ), Age, the Are you a student in paris, the Level of education, the Gender of the participant, the Academic Background and the Culture.

The variables were the Audio as coming from a student (R ), the Interest created by the speaker’s presence, the Competence evaluation and the Audio as coming from a professional (R ). For significance tests, a t(32) “t-test” was used for each predictor, and the resultant t statistic is expressed as the absolute value. In addition, the Tables display the results of R2, the collinearity statistics, and

The results table can be found in the table 3:

**Table 3:**

https://docs.google.com/spreadsheets/d/1jl2UxiLgSHvixneplyCyUj45jmQEQtU6sWRyHWIQCEI/edit?usp=sharing

**Supplementary materials on the JASP output:**

Link to Table 4- Audio as coming from a student (r):

https://docs.google.com/spreadsheets/d/1U9ITlv--gqEtRmSL13RpdhjHBbVdWdGBNVtyfFheGF4/edit?usp=sharing

Link to Table 5- Interest created by the speaker’s presence:

https://docs.google.com/spreadsheets/d/1FMnbgSsUp6xR8SBW\_dyTnFvxoszbIvvjK3E2FMB3uXw/edit?usp=sharing

Link to Table 6- Competence evaluation:

https://docs.google.com/spreadsheets/d/1EIo\_e7qXLHJjqBKmINDkH57iQCtKdXXuZCQMnR7A8X4/edit?usp=sharing

Link to Table 7- Audio as coming from a professional (r):

https://docs.google.com/spreadsheets/d/1jAjGNzb62yiQsfsGOCCGH7JXzsuFT1-IvJvUQUUzp2s/edit?usp=sharing

For the “rating of the audio as coming from a student” variable, related to the following question: “Imagine you are a lecturer at a famous university. For YOUR STUDENTS' final assignment, you decide to ask them to produce a science communication podcast. Your student presents you WITH the recording you JUST heard. WHAT grade would you give them, knowing that A is the best grade, and E is the worst one?”, the results of the regression analysis indicated that none of the gender related listed variables (Gender of the speaker and Gender of the participant) were significantly related to the this variable (Gender of the speaker: t(32)=1.516, p=0.146 > 0.05, and Gender of the participant: t(32)=0.359, p=0.723 >0.05). But the Interest created by the speaker's presence turned out to be a strong predictor of the rating of the audio as coming from a student (t(32): 2.951, p=0.008): everytime the rating of the audio as being a student’s work increases by 1, and if all the other observed factors are maintained constant, the rating of the interest created by the speaker’s presence increases of 0.354 (Unstandardized: 0.354). When looking at the observed covariates, those factors explain 50,4% of the variance of the audio as coming from a student. Finally, the associated F test is insignificant (F(10,21)= 1.612, p=0.171). Therefore, we cannot reject the null hypothesis of an absence of link between the ratings to the cited question and all the covariates. When looking at the partial correlations, 50,7344% of the variance is not explained by our covariates, and 31,4721% of the variance is explained by the ratings of the Interest created by the speaker’s presence.

For the question “Below is a linear scale ranging from 1 (Not at all), to 7 (Very much). To what degree has the speaker interested you in the topic discussed in the audio?”, the results of the regression analysis indicated that none of the gender related listed variables (Gender of the speaker and Gender of the participant) were significantly related to the Interest created by the speaker’s presence variable (Gender of the speaker : t(32)=-0.167,p=0.869 >0.05, and Gender of the participant: t(32)= -1.068, p=0.299 > 0.05). But this regression analysis confirms that the ratings of the audio as a student’s work is a strong predictor of the Interest created by the speaker’s presence (t(32)=2.951, p=0.008): everytime the rating of the interest created by the speaker’s presence increases by 1, and if all the other observed factors are maintained constant, the rating of the audio as being increases by 0.887 (Unstandardized: 0.887). When looking at all the observed covariances, those factors explain 64,9% of the variance of the Interest created by the speaker’s presence. Finally, the associated F test is significant (F(10,21)= 2.932, p=0.018). Therefore, we are allowed to reject the null hypothesis of an absence of link between the answer to the following question and all the covariates. When looking at the partial correlations, 38,1473% of the variance is not explained by our covariates, and 31,4721% of the variance is explained by the ratings of the audio as being a student’s work.

For the question “Below is a linear scale ranging from 1 (not competent) to 7 (extremely competent). How would you categorize the speaker's competence in the topic discussed on the audio?”, the results of the regression analysis indicated that none of the gender related listed variables (Gender of the speaker and Gender of the participant) were significantly related to the speaker’s competency evaluation (Gender of the speaker : t(32)=-0.129, p=0.899 >0.05, and Gender of the participant: t(32)= 0.567, p=0.577 > 0.05). None of the other covariates were significantly related to the named question. When looking at all the observed covariances, those factors explain 42,3% of the variance of the speaker’s competence evaluation. Finally, the associated F test is insignificant (F(10,21)= 1.162, p=0.373). Therefore, we cannot reject the null hypothesis of an absence of link between the ratings to this question and all the covariates. When looking at the partial correlations, 14,9769% of the variance is explained by the answer to the feeling like understood content by the participant, 15,6816% of the variance is explained by the fact of knowing the topic before or not, 12,1801% of the variance is explained by the culture of the participant and 15,6816% of the variance is explained by the academic background.

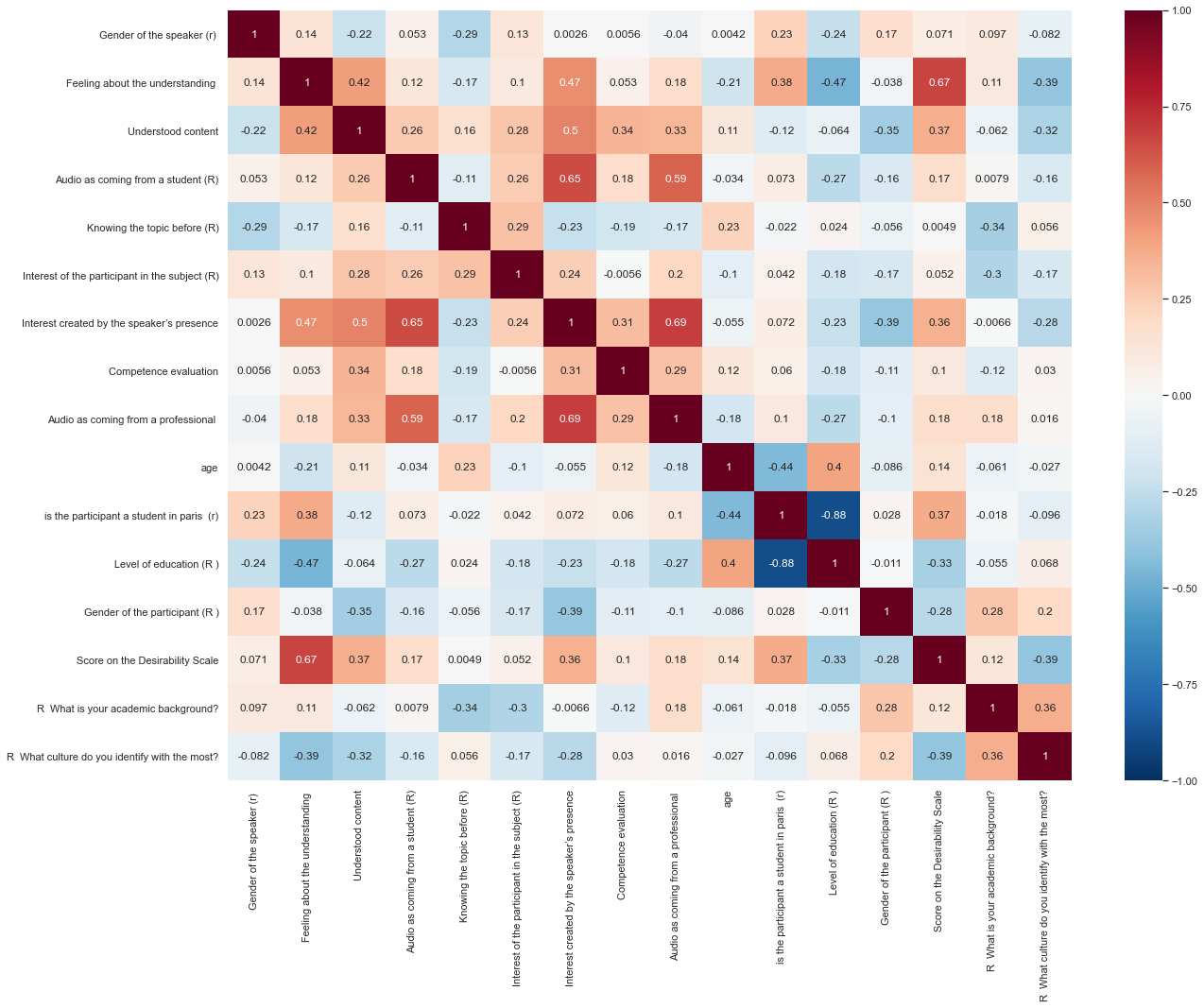
For the question: “Below is a linear scale ranging from 1 (not likely) to 7 (highly likely). You are the founder of a tech start-up in Silicon Valley. You have advertised 2 open positions in your physics team on LinkedIn. In the advertisement you state that the individuals interested in applying for a job should email a self-produced, educational, and entertaining physics video. If you had received an email with the audio you heard previously, how likely is it for you to invite the speaker to an interview?”, the results of the regression analysis indicated that none of the gender related listed variables (Gender of the speaker and Gender of the participant) were significantly related to the speaker’s competency evaluation (Gender of the speaker: t(32)=-0.445, p=0.661 >0.05, and Gender of the participant: t(32)= 0.833, p=0.415 > 0.05). But this regression analysis confirms that the rating of the interest created by the speaker’s presence is a strong predictor of the ratings of this audio as being a professional work(t(32)=3.520 , p=0.002): every time the rating to the precedent question increases by 1, and if all the other observed factors are maintained constant, the rating of the interest into the subject created by the speaker’s presence increases by 0.727 (Unstandardized: 0.727). When looking at all the observed covariances, those factors explain 58,6% of the variance of this question. Finally, the associated F test is insignificant but almost significant (F(10,21)= 2.239, p=0.056). Therefore, we cannot reject the null hypothesis of an absence of link between the ratings to the audio as coming from a professional work and all the covariates. When looking at the partial correlations, 46,7294% of the variance is explained by unknown covariates: 39,4384% of the variance is explained by the variable concerning the interest created by the speaker’s presence. We further explore the implications of these effects in the Discussion section.

All of the collinearity statistics show strong multicollinearity rates (tolerance rates are > 0.1, and VIF are <10). This means that the different model predictors show strong correlations among them, creating high redundancy between covariates and probably measuring the same aspects multiple times. We further investigate multicollinearity through a correlation analysis.

**Correlation analysis**

**Illustration of the correlation analysis:**

**Heatmap 1**. Correlation heatmap for 16 variables.



Next to the discovery of the multicollinearity rates, the correlation analysis was based on 16 variables, linked to questionnaire’s questions (see Table 1) :

* Gender of the speaker
* Feeling about the understanding of the topic
* The amount of actual understood content
* Rating of the audio as coming from a student’s work
* Knowing the topic before
* Interest of the participant in the subject
* Interest into the subject created by the speaker’s presence
* Competence evaluation of the speaker
* Rating of the audio as coming from a professional work
* Age of the participant
* Is the participant a student in paris
* Level of education of the participant
* Gender of the participant
* Score on the desirability scale
* Academic background of the participant
* Culture of the participant

They were entered simultaneously into a correlation matrix in order to predict the impact of each variable on the other one, by pairs.

The results of the analysis are illustrated in the Heatmap 1. It shows Pearson's correlation coefficient value for each pair of variables, which represents the strength of the link between the two variables. In order to find out if the correlation is significant or not, we used additional t-test analysis, which results can be seen in Table 8. This test is being automatically made by JASP, according to multiple statistical criterias. Each correlation analysis is made for a pair of variables, independently of the others.

**Table 8:**

https://docs.google.com/spreadsheets/d/1tU3qNlM4Qk2H4bzLB1ae3M9ivKfp2z0GrVCceYOJ3lY/edit?usp=sharing

The results of the correlation analysis indicated that one of the gender related variables, speaker’s Gender and Participant’s gender, was significantly correlated with any of the other variables (all other p-values >.05). The correlation coefficient showed that the Interest into the subject created by the speaker’s presence, one aspect of the speaker’s competency evaluation, is significantly positively correlated with the gender of the participant (Pearson’s r=-0.39, p<.05). Therefore, we can assume a weak link between those two variables: if the gender of the participant increases with 1, meaning the participant’s gender is female, the interest rated as explained by the speaker decreased by 0.39, and vice versa.

But the correlation analysis confirmed multiple links between other variables. Please note that the correlation coefficient between two variables may be understood in one variable influencing another, but as the second variable influencing the first one two: there is no causal analysis, no direction of the influence. Therefore, the correlation coefficients will be cited only once by pair.

The results of the correlation analysis outside gender indicated that the feeling about the understanding of the topic (Pearson’s r= 0.47, p = .038), the amount of actual understood content (Pearson’s r= 0.5, p=.004), and the score on the desirability scale (Pearson’s r=0.36, p=.045 ) were significantly correlated to the rating of the interest created by the speaker’s presence.

The interest created by the speaker’s presence is one of a triade of very strongly linked variables, with the ratings of the audio as coming from a student’s work and the ratings of the audio as coming from a professional’s work. The correlation coefficient between the interest created by the speaker’s presence and the ratings of the audio as coming from a student’s work is 0.65 (p<.001), the correlation coefficient with the ratings of the audio as coming from a professional’s work is 0.69 (p<.001), and the correlation coefficient between the audio as coming from a student’s work and from a professional’s work is 0.59 (p<.001). This means that the rating to one of these questions is strongly linked to the rating that will be provided to those other two questions, meaning that those questions may be evaluating the same criteria or that the measured factors are strongly linked together.

The results of the correlation analysis outside gender indicated that the feeling about the understood content (Pearson’s r=0.67, p<.01), the amount of actual understood content (Pearson’s r=0.37, p=.04), the speaker’s culture (Pearson’s r=-0.39, p=.026), the speaker being a parisian student (Pearson’s r= 0.37, p=0.038) were significantly correlated to the score on the desirability scale

.

The results of the correlation analysis outside gender indicated that the feeling about the understanding (Pearson’s r=-0.47, p=.038), the speaker not being a parisian student (Pearson’s r=-0.88, p<.001 ) and the age of the participant (Pearson’s r=0.4, p=.024) were significantly correlated to the level of education of the participant.

The results of the correlation analysis outside gender indicated that the participant’s culture (Pearson’s r= 0.36, p=.043) was significantly weakly correlated with the academic background.

The results of the correlation analysis outside gender indicated that the student not being a parisian student (Pearson’s r=-0.44, p=.012 ) was significantly weakly correlated with the age of the participant.

We further explore the implications of these effects in the Discussion section.

**Content analysis**

The content analysis was based on the Open Question: “The person you heard presenting is going to an interview into the university you're studying at to become professor. If you know you're taking their course next semester, would you like them to be hired and why ? “.

We splitted the participant into 6 groups depending on their speaker’s gender and if the participant was in favor, not if favor or did not know if he/she would hire the speaker. The analysis consisted of the counting of the most common words in the answers in each group. The results to this analysis are shown in Table 8, with the descriptive statistics, and in Table 9, with the most common words beside stopwords in each group, the number of appearances, the percentage of appearance of each word and the stemmer.

**Table 8:** Descriptive analysis of the repartition of the subjects on the Open Question

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Group/ Opinion** | **Yes** | **No** | **Don't know** | **Total** |
| **A** | 11 (37,93%) | 2 (6,89%) | 1 (3,44%) | 14 (48,27%) |
| **B** | 5 (17,24%) | 2 (6,89%) | 8 (27,58%) | 15 (51,72%) |
| **Total** | 16 (55,17%) | 4 (13,79%) | 9 (31,03%) | 29 |

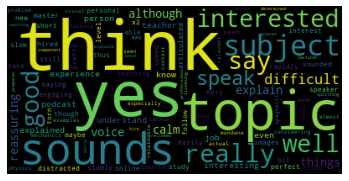
**Table 9** : Please note the entitling of the variables and the visuals may have been given a new version

<https://docs.google.com/spreadsheets/d/1P_39XG6nBjU6OQsTtQBfg5MnQBAcZ7SY-JNZ-79L-dc/edit?usp=sharing>

In the group A (male speaker), 37,93% of the participants answered they would hire the speaker (Group Ayes), for 6,89% that would not hire the speaker (Group Ano), and 3,44% that answered they did not take a decision (Group Adontknow).

The participants that answered they would hire the speaker showed very positive words stems when looking at the most common words among their answers: interest (0,16% of the words are “interested”), studi (0,05% of the words are “studying”), learn (0,05% of the words are “learning”), think (0,10% of the words are “think”), explain (0,10% of the words are “explained”), reassure (0,10% of the words are “reassuring”, 0,10% of the words are “calm”), etc.

**Wordcloud** of the most common words for Open Question1. Group Ayes



The participants that answered they would not hire the speaker showed negative words stems such as monoton (1,16% of the words are “monotone”) or neutral words stemmer such as person (1,16% of the words are “person”).

**Wordcloud** of the most common words for Open Question 2. Group Ano



The participants that answered they made no decision about hiring the speaker or not did show positive words stems such as explain (0,5% of the words were “explanations”), scientific (0,5% of the words were “explanations”),correct (0,5% of the words were “correct”) and understand (0,5% of the words were “understand”), doubt-related words stems such as would(0,5% of the words were “would”) and seem (0,5% of the words were “seems”), but they did write about drawing being useful: draw(0,5% of the words were “drawing”), etc.

**Wordcloud** of the most common words for Open Questions3. Group Adontknow



In the group B (female speaker), 27,58% of the participants answered they did not take a decision regarding the hiring of the speaker (Group Bdontknow), for 17,24% that answered positively to the hiring decision (Group Byes), and 6,89% that would not hire the female speaker (Group Bno).

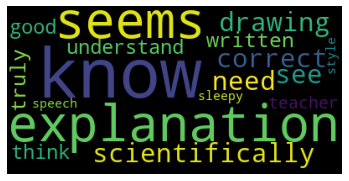
The participants that answered they would hire the speaker showed neutral words stems when looking at the most common words among their answers: teach (0,09% of the words are “teaching”), class (0,19% of the words are “class”), explain (0,09% of the words are “explained''), science (0,09% of the words are “science”), physic (0,19% of the words are “physics”), etc. And they manifested the use of visuals (0,29% of the words were “visual”).

**Wordcloud** of the most common words for Open Question 4. Group Byes



The participants that answered they would not hire the speaker showed doubting words stems such as seem(0,16% of the words are “monotone”), would (0,45% of the words are “would”), think (0,45% of the words are “think”), negative words stems such as spleepi (0,45% of the words are “sleepy”) but positive words stems two, such as correct (0,45% of the words are “correct”).

**Wordcloud** of the most common words for Open Question 5. Group Bno



The participants that answered they made no decision about hiring the speaker or not mentioned the visuals two (0,10% of the words were “visual”), but most of the words stems used are related to calling the speech being not enough to judge: enough (0,16% of the words were “enough”), “time” (0,05% of the words were “time”), minut (0,05% of the words were “minutes”), condens (0,05% of the words were “condense”), introduct (0,05% of the words were “introduction”), short-term (0,05% of the words were “short-term” and 0,05% of the words were “short”).

**Wordcloud** of the most common words for Open Question6. Group Bdontknow



**Link to raw datas:** https://docs.google.com/spreadsheets/d/1I0s6xKWiY3iK1bUykYxTY5sBovwWTStiPkYCm3W-F4Q/edit?usp=sharing

## **Discussion**

The 5 key results:

* (regression analysis) None of the gender related factors (gender of the speaker or gender of the participant) are predictors of the speaker’s evaluation. Therefore, we cannot confirm our main hypothesis about the presence of gender stereotypes in our study.
* (correlation analysis) It seems like female participants rated the following question : “To what degree has the speaker interested you in the topic discussed in the audio?” lower than the male participants. This shows a small link between the participant’s gender and its ratings of the speaker’s ability to interest the participant into this subject.
* There is a strong link between three criterias: the question about the speaker’s ability to interest the participant into this subject, the ratings of the audio as being a student’s work and the ratings of the audio as being a professional’s work. This can be understood as either all of the questionnaire’s questions measure the same point, or either each of the three measured factors are strongly linked together (meaning a rating given on one of these questions can predict the rating given on the other two). This is not enough to validate our hypothesis, but this is still a noticeable point.
* (Content analysis) Participants tend to hesitate more about hiring the female than the male speaker, and tend to hire the male speaker more then the woman speaker. Even though all speaker’s gender combined, participants report that visuals would have been useful.
* (Content analysis) First, participants justify their hesitation to hire the female speaker by reporting a major absence of sufficient content to judge her competence, while the hesitation to hire the male speaker baths in positive remarks. Second, participants tend to have a more positive attitude towards the hiring of the male speaker, and presenting a more neutral approach when discussing the hiring of the female speaker.

An interesting finding from the multiple statistical analysis is that female participants tend to rate the speakers teaching proficiency more positively than their male counterparts, even when they held negative views towards them. They tended to use more positive words and demonstrated a more forgiving attitude to both the speaker's tone and enunciation. This phenomenon was exhibited in both groups A and B. It is suggested that the gender of the speaker was not considered a significant factor in assessing their characters for female participants.

The differences in gendered prosocial behavior have been widely studied. Some of the literature suggests this stems from differing neurochemistry between genders as the primary determinant for the difference in behavior (Tobler PN. et al 2017). It is a widely held belief that females are less competitive in general. Thus, they are less likely to engage in selfish behavior than males (Eckel & Grossman. 1998). The 1998 study highlighted that females were significantly more likely to engage in altruistic behavior than males, which might have some grounding in the results that this study obtained. On the other hand, as illustrated by the Bobo doll experiment, children readily adopt the behavior of their elders if there are certain perceived societal inclinations toward those behaviors (Nolen JL. 2020). This behavior is further supported by the presence of behavioral imprinting observed in animals. During a certain phase of a young animal's life, the behaviors of an adult are repeated by the child regardless of consequences (McCabe BJ. 2019). Although the reason behind the tendency for females to respond with more positively inclined words to the survey in this study is likely a mix of several factors, it does present itself often in literature. To which degree this is determined by neurochemistry rather than societal priming remains up for debate and is not covered under the scope of this study.

Females in this study reported being less interested in the topic than their male counterparts (Table 8). Following from the result of the correlation analysis, females are not only correlated as being less interested in the topic, but also that the gender of the speaker had no noticeable impact on their level of interest. On the other hand, while the male participants' interest in the topic remained unaffected by the gender of the speaker as well, they were, as a group, significantly more interested in the topic. If we assume a social construct viewpoint, this result can somewhat be understood through the Steele stereotype-threat hypothesis (STH) (Osborne JW, 2001 & Steele CM. 1997). STH attempts to illustrate that group stereotypes can affect the performance of individuals in an academic setting by enforcing differential expectations onto the individuals. Steele explains that academic achievement and performance is influenced by the anxiety of performing according to the group stereotypes. However, this is contrasted by the gender-equality paradox (GEP), which suggests a strong inverse correlation between the Global Gender Gap (GGG) Index of individual countries, and the gendered differential preferences of its citizens (Geary & Stoet. 2018, Giolla & Kajonius. 2018). It is illustrated that the more egalitarian a country becomes, the wider the gender gap in preferences and personalities seem to get. It is also suggested that stereotypes and gender roles are not the primary determinant for this gap, but rather endogenous differences between males and females. France has consistently ranked in the top 20 countries worldwide (World Economic Forum Global Gender Gap Report 2020) on the GGG Index, which may lend some credence to the GEP, and therefore might serve as an explanation of the gendered differences in responses to the participants interest in the topic.

This study included a total of 32 participants, which is insufficient to achieve significant results (Linda Kamas et al. 2012). Perhaps with more participants, the analyses would have shown more effectiveness for assimilating the insight of the people. Further analysis of some questions has shown that there are different perceptions of participants towards the genders. This approach has been attained through qualitative content analysis. The results have revealed the indirect gender differences. Besides, some unconscious differences of judgments have been immersed from the analysis of results. However, the applied method seems to unable to present the real scenario of gender discrimination. For better understanding, further research needs to be conducted with the precise methods and questionnaire.

Other possible explanations may be that the questionnaire psychometric qualities were not investigated, making the questionnaire possibly not usable nor pertinent. Another possibility would be that the audio without any visual informations did not generate enough gender discrimination for our study to uncover them.

## **Speculation**

**Self-reflection about the research process**

About the research process, the more challenging things for the team when developing your research project were the idea building, the creation of the experiment materials (audios, questionnaire) from scratch, collecting enough data and finding ways to spread the study, the writing of clear results and analyzing them to extract a meaningful discussion, the small amount of data and extraction of the results, the linking of everyone’s interests together into finding a unique subject, the communication/online meetings disturbing the team building, and the competence of staying motivated.

We experienced openness and collaboration during the process by using a free open source software (JASP) for the results, and providing those results, as well as the raw answers to the questionnaire (anonymized). We worked together at every step of the way, even though the team was made from multiple academic and cultural backgrounds, our sample was diverse, and we collaborated to include everyone’s ideas.

Concerning future research, if we could redo our study from scratch, we would change a few things. We would take more time to recruit more participants and use a budget for the recruitment, take more time to work and review the article, we could use more regular feedbacks two, and we would focus on the creation of the questionnaire more than on the results that were not significant. We would take time to make a pre-questionnaire construction with open meetings with participants two, and measure the psychometric qualities of our questionnaire before using it.

X= the person was implicated

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Names** | **Lisa** | **Imran** | **Marine** | **Farah** | **Carlos** | **Kane** |
| **Conceptualization** | X | X | X | X | X | X |
| **Data curation** | X | X | X | X | X | X |
| **Formal analysis** | X | X | X | X | X | X |
| **Investigation** | X | X | X | X | X | X |
| **Methodology** | X | X | X | X | X | X |
| **Software** | X | X | X | X | X | X |
| **Visualization** | X | X | X | X | X | X |
| **Figures** | X | X | X | X | X | X |
| **Writing—original draft** | X | X | X | X | X | X |
| **Writing—review and editing** | X | X | X | X | X | X |
| **Project administration** | X | X | X | X | X | X |

**References:**

Bandura, A., Ross, D., & Ross, S. A. (1963). [Imitation of film-mediated aggressive models](https://www.uky.edu/~eushe2/Bandura/Bandura1963JASP.pdf). *The Journal of Abnormal and Social Psychology*, 66(1), 3.

Eckel, C. & Grossman, P. (1998). Are Women Less Selfish Than Men?: Evidence From Dictator Experiments, The Economic Journal, 108(448), 726–735, https://doi.org/10.1111/1468-0297.00311

Geary, D. and Stoet, G. (2018) The Gender equality paradox in science, technology, engineering and mathematics education. Pubmed, 29(4). DOI: [10.1177/0956797617741719](https://doi.org/10.1177%2F0956797617741719). Found at: <https://journals.sagepub.com/doi/abs/10.1177/0956797617741719>

Haghighat, R. (2013). The Development of the Brief Social Desirability Scale (BSDS). Found at:

<https://ejop.psychopen.eu/index.php/ejop/article/view/417/html#:~:text=The%20assumption%20underlying%20the%20construction,the%20predisposition%20to%20stigmatize%20others>.

Hofstede, G. (1986). Cultural differences in teaching and learning. International Journal of Intercultural Relations, 10(3), pp 301-320. DOI: [10.1016/0147-1767(86)90015-5](https://doi.org/10.1016/0147-1767(86)90015-5). Found at: https://www.sciencedirect.com/science/article/abs/pii/0147176786900155

Huan He. (2018) A case study of the effects of geographical teachers‘ sex on classroom teaching behavior. Jiangxi Normal University, MA thesis.

Jepsen, D. M., & Rodwell, J. J. (2009). Justice in the workplace: The centrality of social versus judgmental predictors of performance varies by gender. The International Journal of Human Resource Management, 20(10), 2066-2083. https://doi.org/10.1080/09585190903175639

Jones, M., Howe, A. and Rua, M. (2000). Gender differences in students’ experiences, interests, and attitudes toward science and scientists. Science Education, 84(2). DOI: [10.1002/(SICI)1098-237X(200003)84:2<180::AID-SCE3>3.0.CO;2-X](https://doi.org/10.1002/(SICI)1098-237X(200003)84:2%3C180::AID-SCE3%3E3.0.CO;2-X). Found at <https://onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1098-237X(200003)84:2%3C180::AID-SCE3%3E3.0.CO;2-X>

Kamas, L. & Preston, A. (2012). The importance of being confident; gender, career choice, and willingness to compete. Journal of Economic Behavior & Organization, 83(1), 82-97. https://doi.org/10.1016/j.jebo.2011.06.013

Kenton, W. (2020). T-Test. Found at: https://www.investopedia.com/terms/t/t-test.asp#:~:text=A%20t%2Dtest%20is%20a,be%20related%20in%20certain%20features.&text=A%20t%2Dtest%20looks%20at,to%20determine%20the%20statistical%20significance.

McCabe BJ (2019) Visual Imprinting in Birds: Behavior, Models, and Neural Mechanisms. Front. Physiol. 10:658. doi: 10.3389/fphys.2019.00658

Nguyen, L.T , Nguyen H. T. & Pham, M. H. (2020) Bribe or die: gender differences in entrepreneurship in emerging markets. Applied Economics. <https://doi.org/10.1080/00036846.2020.1856325>

Leyens, J-P. (1997) Psychologie Sociale. Mardaga.

Lin Yao. A study on gender role and teaching style of primary school students. 2018. Qufu Normal University, MA thesis.

Lixia Wang. A gender-based study of teachers’ questioning in junior high school mathematics class. 2019. Tiansui Normal University, MA thesis.

Jingyuan Chai. Research on the relationship between the gender role types of junior middle school teachers and student evaluation. 2020. Southwest University, MA thesis.

Pla-Julian, I. and Diez, J-L. (2018) Gender equality perceptions of future engineers. Engineering Studies, 11(3), pp 243-251. DOI: [19378629.2018.1530242](https://doi.org/10.1080/19378629.2018.1530242). Found at: <https://www.tandfonline.com/doi/full/10.1080/19378629.2018.1530242?scroll=top&needAccess=true>

Xiaofang Li. Research on the difference of the response of male and female teachers to primary school pupils‘s different behaviors. 2018. Gannan Normal University, MA thesis.

Junming Huang et al., “Historical Comparison of Gender Inequality in Scientific Careers across Countries and Disciplines,” *Proceedings of the National Academy of Sciences* 117, no. 9 (March 3, 2020): 4609–16,<https://doi.org/10.1073/pnas.1914221117>;

“Schools’ Racial Mix, Students’ Optimism, and the Black-White and Latino-White Achievement Gaps - Pat António Goldsmith, 2004,” accessed January 26, 2021,<https://journals.sagepub.com/doi/10.1177/003804070407700202>;

Jason W. Osborne, “Testing Stereotype Threat: Does Anxiety Explain Race and Sex Differences in Achievement?,” *Contemporary Educational Psychology* 26, no. 3 (July 1, 2001): 291–310,<https://doi.org/10.1006/ceps.2000.1052>;