

Experimental Designs in Management Studies for Bias Assessment: Setting and Stimulus Material

Maria João Guedes, Ricardo Rodrigues and Luís Gouveia

ADVANCE, Centre for Advanced Research in Management – ISEG, ULisboa, Portugal

mjguedes@iseg.ulisboa.pt

rmrodrigues@iseg.ulisboa.pt

luisgouveia@iseg.ulisboa.pt

Abstract: This manuscript stems from an ongoing research project on gender inequality in Portuguese businesses, particularly concerning access to leadership positions. It aims to provide new insights into the constraints that limit women's access to managerial roles – both from the worker/employee's perspective, by exploring expectations, aspirations, and perceived barriers, and from the employer/hierarchy's perspective, by analysing which biases in recruitment and promotion processes are likely to limit women's rise to top management positions. Against this analytical background, this paper aims to provide a reflection on the inquiry protocol developed within the project, with particular emphasis on the quantitative approach based on an experimental design. A survey has been developed targeting professionals and directors of large, listed companies and large/medium unlisted companies, where the respondents are given the framework of hiring hypothetical applicants for a management position. The methodological protocol involves elaborating different fictitious CV profiles, representing hypothetical candidates. These profiles are constructed by varying attributes according to various dimensions, specifically gender, combined with educational and professional characteristics. Based on the process of designing the experimental setting, the aim is to reflect on the nuances, obstacles, and limitations in constructing a set of stimulus materials and manipulating the experimental conditions regarding the fictitious applicants' characteristics (including gender), vertical segregation ('glass ceiling' in access to top management positions), horizontal segregation (i.e., associating candidates with typically female or male occupations) and measurement items/scales for assessing the applicants in different dimensions (e.g., perceived competence, job fit, promotion recommendation, etc.). Therefore, this paper aims to reflect on the methodological challenges and advantages of an experimental research design and data analysis strategy to address gender bias in hiring contexts and capable of sustaining an ongoing data observatory to support the definition of public policies and organisational practices aimed at promoting gender equality in top positions.

Keywords: Quantitative methodology, Management studies, Experimental survey, Bias assessment

1. Introduction

The issue of women's inequality in businesses, particularly in top management positions, has been a topic of interest among scholars, practitioners, policymakers, and the media (Terjesen et al., 2009). The debate is mainly centred on utility arguments (based on business cases) and justice rationales (Seierstad, 2006; Terjesen & Sealy, 2016).

In the specific case of Portugal, new legislation introduced in 2017 established that the management and supervisory bodies of public sector companies and listed companies should meet a 33% threshold for the underrepresented gender. Notwithstanding this legislative effort, statistics confirm the inequalities persist, with women receiving, on average, less 14.4% than men for each month of work. This number is further aggravated to 17.8% if one includes other forms of remuneration, such as bonuses or other discretionary compensation items. Furthermore, the average proportion of women on the boards of directors in the listed firms is only 22% (Casaca et al., 2020). The causes of this inequality have yet to be fully explored.

This is a fertile research field, and several advances have been made. One stream of studies focuses on the impact of the presence of women in companies concerning outcomes such as performance, innovation capabilities, etc. (e.g., Adams et al., 2012). Another research group focuses on the factors that explain the differences in representation between men and women in companies, specifically in top management positions (e.g., Eagly et al., 2019).

Along with the different theoretical perspectives, different methodological strategies have been developed. Particularly, research on gender-related topics and boardrooms is based on cross-section secondary data (Malik & Makhdoom, 2016) and scarcely on longitudinal data on firms (Dezsö & Ross, 2012). Using longitudinal data helps address some of the existing research gaps. However, as databases are not specific to address gender research questions, there will always be several unanswered research gaps.

This manuscript is part of a research project whose methodological protocol departs from the more conventional approaches to gender inequality based on secondary data and longitudinal data obtained from companies, but not specifically geared towards the gender issue, i.e., not having the heuristic potential associated with data

collection that is specifically oriented towards the subject in question. In particular, the objective is to apply a survey of experimental type, aiming to address the existing data needs while simultaneously contributing to answer some of the most common criticisms of academic research (Gastaldi & Corso, 2014), including the insufficient knowledge about the reality (from the practitioners' perspective) of the subject under analysis and not being able to transform the knowledge in real benefits for organizations (knowledge transfer).

Experimental studies, with a particular focus on identifying biases in organizational and hiring contexts, can be divided into two main types: *field experiments*, where bias is assessed in real-world hiring decisions involving responding to job vacancies (e.g., Birkelund et al., 2022) and *laboratory experiments*, where participants evaluate hypothetical/fictitious candidates within a setting previously defined in terms of hiring context, job position, etc. (e.g., Carlsson et al., 2021; Dutz et al., 2023; Kessler et al., 2019). It is precisely in the field of laboratory experiments that the survey was developed within this research project focusing on gender bias. The theoretical framework and research hypotheses guiding the methodological protocol and the experimental setting are presented below.

2. Theoretical Background and Research Hypotheses

The extensive scientific covers different disciplinary areas exploring the inequality and discrimination against women in the workplace. The *glass ceiling* consists of a dynamic of *vertical segregation* that hinders women from reaching top management positions and taking higher responsibilities in the corporations/organisations' hierarchy (Eagly et al., 2019). The first barrier identified in women's access to top management positions in organisations concerns family responsibilities, namely the reconciliation of personal and professional life. Indeed, a family project can be a stigmatising factor, associated with less commitment to the professional career and consequent loss of opportunities for advancement. Furthermore, this circumstance can be exacerbated by the lack of flexible working policies in companies/organisations – like flexible working hours, part-time schedules, and telework (Munsch, 2016).

Another central factor is associated with socially constructed gender stereotypes, which have an impact on the attribution of leadership characteristics and competencies. Men can be perceived as having professional and social skills more suited to the role of leader, namely *agentic characteristics*, seen as assertiveness, forcefulness, and ambition – unlike women, who are ascribed *communal characteristics*, such as sensitivity, caring, kindness, etc. (Tremmel & Wahl, 2023). These gender roles socially constructed can favour the undervaluation of women when it comes to accessing top management positions, as well as negative reactions when it comes to exercising leadership positions (Dutz et al., 2023).

In addition to *vertical segregation*, *horizontal segregation* occurs when gender distribution is disproportionate to job functions, i.e., when one gender prevails over the other in a particular job (Eagly et al., 2019). To that extent, women's status in the labour market can be an advantage or disadvantage, depending on the dominant occupational stereotype – male or female. However, the literature suggests that discrimination against women is accentuated when they apply for typically male professions, which is related to the socially prescribed stereotypes (MacNeill et al., 2014).

Considering this theoretical framework on gender bias, we aim to ascertain the existence of discrimination based on gender through an experimental survey focused on controlled manipulation of the content of fictitious professional CVs – as the primary mean for employers to infer qualities and competencies of applicants for hiring purposes – in a context of hiring for a management position (in this case, middle management). Based on the literature on gender stereotyping, as well as vertical and horizontal segregation, the following research hypotheses were formulated:

A) Sex effect

H1: Women are evaluated as less competent and less hireable, for management positions, compared to male candidates with the same qualifications and professional experience;

H2: Women are more overlooked when it comes to hiring for a management position than equally qualified male candidates.

B) Sex effect and horizontal segregation:

H3: Female candidates are evaluated as less hireable to management positions in dominant male professions/areas than equally qualified male applicants;

H4: Female candidates are evaluated as more promotable to senior management positions in typically female areas than equally qualified male applicants;

H5: Female candidates are evaluated as less promotable to senior management positions in typically male areas than equally qualified male applicants.

C) *Sex effect and vertical segregation:*

H6: Female candidates are evaluated as less promotable to senior management positions than male candidates with the same qualifications and professional experience.

Based on the assumption that, in the absence of discrimination, qualified people with the same qualifications can be expected to have the same evaluations and proportion of favourable hiring decisions, regardless of their sex, the experimental setup was developed to test the hypotheses listed.

3. Methods

3.1 Experimental Setting

The present research aims to analyse the circumstances/dimensions under which discrimination against women in access to leadership positions may occur: (i) whether bias occurs in the assessment of competence and hireability between equally qualified job applicants of different sexes in access to management/leadership positions; (ii) whether discrimination is exacerbated in the transition between access to middle management positions and top management positions; (iii) whether discrimination against women intensifies in typically male positions/areas.

Stemming from the guiding research question and the set of hypotheses structuring it, a laboratory-type survey was developed. In laboratory experiments, the methodological design can, firstly, be divided into *within-subjects manipulation*, where participants form a unique group and are subjected to a test with the same content (Dutz et al., 2023; Erlandsson et al., 2023; Barron et al., 2022), and *between-subjects manipulation*, in which participants are distributed between different groups according to controlling variables (Friedmann & Efrat-Treister, 2023; Ramos & Félix, 2019; Hoover et al., 2018). This experimental study is based on the latter, with each participant having access to different combinations of the stimulus material (randomly assigned) to compare responses across different experimental conditions.

The experimental survey encompasses two stages, with different experimental designs that offer laboratory-like controlled conditions. The first stage uses an *unmatched pair-testing* design (Bendick & Nunes, 2012) presenting pairs of CVs of applicants for the same hypothetical job vacancy. Applicants' characteristics – e.g., education, work experience, technical skills, etc. – are differentiated within each pair. These characteristics are experimentally manipulated to detect which features trigger different evaluations from respondents (Bendick & Nunes, 2012). In the second stage, we resort to an *unpaired testing* design (Carlsson et al., 2021), where each respondent is presented with a single CV.

In the first stage of the survey, the setup consists of five experimental conditions. For each experimental condition, two CVs were constructed and set against each other in pairs. These fictitious CVs were structured to ensure the variability of the candidates' attributes in 4 dimensions:

- *Age*. For this parameter, the fictitious candidates vary between three categories: *24/25 years old*, *28/29 years old*, and *41 to 44 years old*;
- *Academic qualifications*. Includes three levels: *low* (Bachelor's degree), *medium* (Bachelor's and Master's degree), and *high* (Bachelor's, Master's and Postgraduate degrees);
- *Type of academic path*. Fictitious applicants with training entirely in the field of *Management* or combining different fields. The latter is divided into two paths: initial training in *Humanities* or *Engineering*, followed in both cases by a Master's in *Management*;
- *Professional experience*. The CVs were built according to four categories: *no experience* (limited to consultancy work in non-profit organisations run by university students); *low* (1-2 internships and 1-2 professional experiences in the labour market over 5 years); *medium* (1-2 internships and 1-2 professional experiences in the labour market over 9-10 years) and *high* (3-4 professional experiences in the labour market over 19-20 years);
- *Marital status*. This includes marital status (married or single) and the number of children (divided into *no children* or *2 children*).

Based on these pre-defined conditions, 10 CVs were created from scratch. With the variation in gender (male vs. female), a total of 20 CVs can be evaluated by the respondents. Moreover, in building the setting for the survey, the position of *Project Manager* was selected as the recruitment context. No reference is made to a specific sector of activity. The aim was not to introduce additional complexity to the setup (i.e., multiplying the experimental conditions) concerning typically male or female areas and its effects on evaluating applicants. Thus, the setup consists of five experimental conditions in which different pairs of CVs are presented sequentially for a comparative evaluation. It should be noted that a randomisation system was also introduced into the questionnaire in each setup. In particular, the gender attached to each fictitious CV varies between respondents.

We proceed with a description of the five experimental conditions presented sequentially to each respondent:

1. The first condition compares CVs with equivalent academic qualifications. Considering the age difference of profile B, the professional experience (*medium*) is slightly higher compared to profile A (a recent graduate with no professional experience);
2. In the second condition, CVs with contrasting qualifications and experience are compared: on the one hand, a CV with high academic qualifications and low experience (profile C); on the other, a CV with a degree but high experience (profile D);
3. In the third setup, the first profile (E) is characterised by a bachelor’s degree in Engineering and a Master’s in Management. In contrast, profile F corresponds to a candidate with a more *conventional* academic path of a project manager, in the sense that it is entirely in the field of Management (Bachelor’s and Master’s degree). The professional experience is equivalent – both in the number of internships and labour contracts;
4. In the fourth setup, a comparison is made between a first profile (G) characterised by a bachelor’s degree in Engineering and a master’s in Management and a second profile (H) characterised by a degree in *Humanities* (Languages, Literature and Culture course), followed by a master’s degree, also in Management. Their professional experience is equivalent;
5. Finally, in the last setup, controlling the candidates' attributes focuses on their family situation. With equivalent academic qualifications and professional experience, the first candidate is single, while the second is married with 2 children.

The following figure summarises the five experimental conditions described (Figure 1).

Experimental condition (nº)	CV (Code)	Age	Civil status	Education	Experience	Setup
1	A	24	Single	Bachelor's and Master's in Management	None	Setup "Age"
	B	29	Single	Bachelor's and Master's in Management	Low	
2	C	25	Single	Bachelor's, Master's and Postgraduate Degrees in Management	Medium	Setup "Academic qualifications vs Experience"
	D	41	Married	Bachelor's in Management	High	
3	E	30	Married	Bachelor's in Engineering and Master's in Management	Medium	Setup "Conventional path v non-conventional path"
	F	29	Single	Bachelor's and Master's in Management	Medium	
4	G	29	Married	Bachelor's in Engineering and Master's in Management	Medium	Setup "Engineering vs Humanities"
	H	30	Single	Bachelor's in Humanities e Master's in Management	Medium	
5	I	45	Single	Bachelor's, Master's and Postgraduate Degrees in Management	High	Setup "Children vs childless"
	J	43	Married, 2 children	Bachelor's, Master's and Postgraduate Degrees in Management	High	

Source: Own elaboration.

Figure 1: Experimental conditions

The aim of this experimental setup is to allow for a comparison of respondents’ evaluations, mapping out how the different variables being analysed – *gender, academic qualifications, type of training, level of professional experience*, etc. – impact the judgements made about the candidates. The purpose is, thus, to identify possible statistically significant discrepancies in the assessment of the same pair of CVs depending on whether the candidate is male or female.

The second stage of the experimental survey is characterised by an *unmatched design*. One of the 20 constructed CVs is randomly distributed among the respondents for a singularised assessment of this fictitious candidate. This different methodological design aims to obtain a comparison of the respondents' assessment of the different profiles built (from A to J), determining whether the evaluations vary according to the sex of the fictitious applicant – as well as according to variables characterising the respondents (sex, age, etc.). Through this overall setup that makes up the experimental enquiry, the aim is to map the types of biases that weigh on the curriculum evaluations made:

- *Gender bias* (i.e., whether women are penalising in the evaluation compared to equally qualified men);
- *Gender bias associated with age* (i.e., whether the level of professional experience is a more penalising factor for women compared to men);
- *Gender bias associated with academic qualifications* (i.e., whether the level of academic qualifications is a more penalising factor for women compared to men);
- *Gender bias associated with the type of academic path* (i.e., whether a conventional/unconventional academic path is a more penalising factor for women compared to men);
- *Gender bias associated with marital status* (i.e., whether the family situation – in particular, parental responsibilities – is a more penalising factor for women compared to men).

Finally, the overall experimental setup is subdivided into different variants that should be distinguished. “Study 1” refers to the survey specifically aimed at university students, with its distinct (and simplified) experimental features. To explore the research hypotheses associated with gender bias (and possible links with the other variables), an additional survey variant was constructed from the more complex experimental setup described. Namely, if the described five experimental conditions compose “Study 2”, in a supplementary survey, the candidates present the same profile, but the sex is the same in each pair. This variant, called “Study 3”, does not include the component relating to isolated CVs. For space reasons of space constraints, the present text focuses on the results from “Study 2” and “Study 3”.

Stimulus Material

The fictitious CVs for the experimental study were created by researching professional profiles and job postings on the professional social network *LinkedIn*. The aim was to achieve verisimilitude in terms of the academic and professional backgrounds of project managers, as well as in the descriptions of professional experience, tasks/responsibilities held, skills, etc. Once the collection of curricular elements was built, these were then used to construct the fictitious CVs according to the described setup parameters.

Before presenting the paired CVs, the same framing text is presented to each respondent:

A multinational consultancy company is looking to recruit professionals for the position of Project Manager to join its expanding team. By the deadline for replying to the job offer, 10 applications have been received.

Imagine that you are a member of the recruitment team at that company. Please consider the paired CVs and for each pair indicate which candidate you would hire as Project Manager.

The presentation of the paired CVs follows this introduction. To ensure effort saving for respondents, a summary version of each CV is presented at this stage, including date of birth, marital status, academic qualifications (years and training institutions) and professional experience (years and companies/organisations).

As for the second stage of the survey, the full version of each CV (with a detailed description of the tasks performed in each work experience, skills, etc.) is randomly assigned. Respondents are presented with the following framing text:

On a scale of 1 to 10, please rate the following full CV of one of the 10 applicants.

3.2 Measures – Outcomes and Other Variables

In the first stage of the experimental survey, three dependent variables are used to evaluate several pairs of fictitious CVs. The first set consists of two items, taken from another experimental survey focusing on gender biases in the hiring process (Kessler et al., 2019): the first measures the respondents' interest in hiring each applicant (“*How interested would you be in hiring the applicants for the position?*”); the second, the perceived likelihood of candidates accepting a position in the respondents' respective company/organisation (“*How likely do you think candidates are to accept a position in your company/organisation?*”). Both variables use 10-point

Likert scales. A third variable concerns choosing which, of the two hypothetical applicants, the respondent would hire for the job ("Which one of these two people would you hire for the job in question?").

In the second stage of the experimental survey, one of the 20 CVs is randomly assigned to each respondent. The evaluation is made using a specific set of dependent variables. Firstly, two indexes are used, both taken from a study on gender bias in recruitment in the academic context (Carlsson et al., 2021).

a) Competence index, made up of three items ("To what extent do you consider the candidate to be competent for the position?"; "To what extent do you consider the candidate to have the necessary qualifications for the position?"; "How qualified do you consider the candidate to be?");

b) Hireability index, made of three items ("How do you assess the probability of the candidate being selected for the position in question?"; "How likely would you invite the candidate for a job interview?"; "Would you hire the candidate in question?").

An additional variable concerns the perception of the likelihood of reaching a top leadership position (in this case, in the company in question) – therefore, gender bias in terms of *vertical segregation*, particularly access to top management positions ("How likely do you think it is that the applicant will, one day, become the head of the multinational company?").

In the case of the survey aimed at students, a set of additional variables was added to assess the phenomena of *horizontal segregation* combined with *vertical segregation*. In other words, the perceived likelihood of an applicant reaching top management positions according to different corporate areas (Chief Financial Officer, Human Resources Director, etc.) – therefore, according to areas characterised by male and female dominance and with possible implications in terms of the privileging of *agentive* or *communal* characteristics in the exercise of leadership (Tremmel & Wahl, 2023).

Finally, the respondents were asked a series of questions for socio-demographic -characterisation:

a) Socio-demographic characterisation: Sex (binary variable, Woman or Man); Age (metric variable); Nationality (binary variable; Portuguese or Other); Marital status (nominal variable); Number of children (if applicable); Academic degree (nominal variable); Years of professional experience (metric variable); Sector of activity (nominal variable with multiple categories, constructed from the CAE, Portuguese Classification of Economic Activities)¹;

b) Structure of daily activities: metric variables to describe the percentage of average daily time spent on each activity (paid work, domestic work, unpaid philanthropic work and leisure activities/hobbies) – both on a typical working day and on a typical weekend day. As gender inequalities in terms of housework and family responsibilities are a factor associated with vertical segregation (Munsch, 2016), these variables are relevant for assessing inequalities in the sample according to gender;

c) Corporate/organisational context: sex of the top manager (binary variable – woman or man); size of the company/organisation (categorical variable); if the respondent has direct responsibilities in terms of human resources in the company/organisation (i.e., recruitment, promotion, etc.).

Other measures were used for three additional dimensions: *professional experience* (e.g., aspirations to access senior managerial positions, expectations concerning career progression, etc.), *gender stereotypes* (concerning leadership skills of men and women), and *political positioning* (namely, about legislative measures to ensure balanced representation between men and women in top management positions).

3.3 Sampling and Data Collection Procedure

The target population of this experimental survey encompasses two main groups: professionals working in organisations in Portugal and university students.

In the case of professionals, while the primary focus is on middle and senior managers, the sample also includes non-managerial employees. In each company/organisation there is the possibility of non-management professionals answering the survey (e.g., a member of a manager's secretarial team). Notwithstanding this methodological obstacle to accessing the primary survey targets (i.e. middle and top managers), respondents in

¹ In the specific case of the survey aimed at students, respondents are also asked about the course they are enrolled in or the higher education institution that they attend.

other positions are also considered, given the analytical relevance of comparing the perspectives of professionals at different levels of the corporate/organisational hierarchy. Regarding the second contingent, we considered the student population from all cycles – Undergraduate, Postgraduate/MBAs, Master and PhD. On a contingency basis, students (from any Portuguese higher education institution) with access to the questionnaire for professionals are redirected to the students' version.

The sampling is *non-probabilistic*, combining the *purposive* and *snowball* methods. A list of companies and email addresses was used to reach the professionals target population. In addition, the professional social network LinkedIn was used as a platform to disseminate the survey. For students, the survey was distributed through the institutional channels of electronic communication between faculty and students – including institutional student email lists and newsletters. As an incentive for participation, respondents are informed at the beginning of the questionnaire that a donation (of €0.5) will be made to a charity for each completed survey. Each respondent chooses the beneficiary institution from a list provided at the end of the survey.

The survey was built, distributed, and administered to participants using the Qualtrics online survey software (www.qualtrics.com).

3.4 Data analysis

Considering the research hypotheses described, the data analysis focuses on identifying discrepancies that indicate gender bias in evaluating CVs. In this context, inferential statistics play a key role in the hypothesis testing. We use independent t-test for dichotomous independent variables (e.g., sex, experience in human resources, etc.) and ANOVA for independent variables with more than two categories (e.g., age categories, academic qualifications, etc.). The main objective is to identify statistically significant differences in the way fictitious applicants are evaluated according to the various parameters defined: existence of gender segregation (*vertical* and *horizontal*), relationships between types of segregation and socio-demographic variables of the respondents, etc.

The inferential statistics are preceded by the validation of the indexes and descriptive statistics – both for the control variables and socio-demographic characterisation, and for the dependent variables. The survey was pre-tested with students to evaluate the construction of the experimental setting and the dependent variables in their effectiveness in identifying significant differences in evaluating candidates.

4. Discussion

Due to space constraints, we focus on the central tendencies, particularly concerning the general values obtained in the different experimental settings for detecting gender bias. Considering the “Study 2”, in the survey block with the direct comparison between CVs of opposite sexes, we observed a tendency to favour the female applicant in all the experimental setups. Namely, across all five arrangements there was a statistically significant increase in the interest in hiring and in the percentage of the decision to hire the female candidate compared to the experimental group in which the same profile (i.e., with exactly the same CV) presents a male applicant. These results are comparable to those obtained in other experimental surveys (Carlsson, 2021; Erlandsson et al., 2023), where there was no evidence of a bias against female applicants. On the contrary, female candidates are perceived as more *competent* and *hireable* than equally qualified male candidates. Furthermore, there was no evidence of a child penalty for either male or female applicants (*Idem*). Other previous studies present statistical patterns that reveal a strong asymmetry in employers' treatment of male and female candidates (e.g., Barron et al., 2022). Namely, when one candidate is more qualified, employers do not display a gender bias in hiring decisions. Still, when CVs are identical, there is a significant gender bias against women in hiring. In the present study, the identification of the project at the beginning of the survey may have played a central role, given that the gender of the candidate was a central variable being analysed.

On the other hand, in “Study 3”, with same-sex pairs of applicants, the results are discrepant. In fact, there is no longer a situation of generalized evidence of gender bias in the form of statistically significant differences in the mean scores for interest in hiring and in the percentages for the hiring decision. The exception is setup 3, where the same female CV (“CV F”) is significantly favoured over the male counterpart when it comes to interest in hiring ($\alpha = 0.025$). However, the same significant differences do not occur in the item concerning the hiring decision ($\alpha = 0.875$).

In the case of the survey block with isolated CVs (thus, a methodologically *unmatched experimental design*), the overall scenario is a more composite one. Particularly, the significant differences are concentrated in the two profiles: the CV characterised by no professional experience and a bachelor's and master's degree in

management (coded “CV A”), and the CV exhibiting the highest academic qualifications (postgraduate degree in addition to a bachelor’s and master’s degrees), extensive work experience, and a marital status indicating that the applicant is *married, with 2 children* (coded “CV J”). Nevertheless, the results diverge on the favoured gender in the items used to assess *competence* and *hireability*.

With regard to the first CV (“CV A”), there are three items in which the difference in the mean scores has a significance level equal to or below 0.05: *How qualified do you think the applicant is* ($\alpha = 0.002$), *How interested would you be in hiring the applicant for the position?* ($\alpha = 0.022$) and *How likely do you think it is that the applicant will be selected for the job?* ($\alpha = 0.003$). In all cases, the male version of the CV obtains the highest values. Moreover, previous studies relying on un-matched design have also identified significant gender differences in the likelihood of being hired favouring male candidates and better assessed for a leadership position in terms of *competence* (Ramos & Félix, 2018).

In the case of the second curriculum (“CV J”), significant differences occur in three items: *To what extent do you consider the applicant competent for the position?* ($\alpha = 0.049$), *How likely do you think it is that the candidate will be selected for the job?* ($\alpha = 0.051$) and *How likely do you think it is that the applicant will one day become the head of the multinational company?* ($\alpha = 0.041$). However, the significant differences in this case favour the female version of the same CV. To this extent, marital status seems to be a more favourable factor in the evaluation of the female CV. This tendency follows previous research valuing the existence of children as a positive element in terms of leadership skills (Erlandsson et al., 2023).

5. Conclusions

The experimental survey presents itself as an alternative to the *field experiment* as a data collection tool for identifying gender bias in organisational contexts. Indeed, measuring unequal treatments in laboratory studies can pose obstacles due to the difficulty in creating realistic situations, as well as assessing whether answers reflect real discrimination and whether employers’ answers are consistent with their behaviour (Birkelund et al., 2022). In turn, the controlled laboratory experiment allows for a more effective systematic variation of environmental aspects while holding others constant (Kessler et al., 2019), thereby opening up opportunities to develop and test different theories concerning the organisational context.

In addition to these advantages in corroborating and refuting research hypotheses and the inherent limitations of this research instrument, it is important to identify specific limitations of the survey developed – concerning precisely the challenge of isolating contributing factors impacting CVs’ assessment. Firstly, concerning the knowledge of the central variable being tested (the applicants’ gender), it is important to consider that the respondents were presented with the title of the project and a brief description. The research team made a methodological decision to adopt a policy of transparency regarding the purposes of the research. In turn, this option may have favoured women candidates more in the block relating to the direct comparison between CVs. Notwithstanding, the isolated CVs, as a supplementary experimental configuration integrating the survey, allowed for a comparison between the mean scores for each profile according to the gender associated. This complex experimental configuration, combining different methodological designs, aimed to neutralise the limitations associated with each type of experimental setup as far as possible – i.e. direct comparison (*matched design*) and isolated assessment (*unmatched design*).

In addition, other potential sources of bias, whose control is challenging, can concern particularly the design of the CVs’. This is the case with different higher education institutions, in the case of academic training, or the greater or lesser reputation of companies composing work experience. The plausibility of the hypothetical CVs forced the diversity of training paths and professional experiences in different organisations. However, the intention to convey realism through this plurality can also result in a lack of control over the influence of contingent factors in how profiles are assessed. An alternative approach adopted in other research consists of more abstract descriptions of qualifications, such as ratings given to applicants’ different types of knowledge (e.g., Barron et al., 2022). However, this methodological option may reinforce the laboratory nature of a survey, thereby compromising the intended immersion of the respondents in a realistic exercise of evaluating job applicants – as closely as possible to a *field experiment*.

Thus, among the multiplicity of methodological options, the present survey sought to differentiate from previous research through an experimental design that combines different modalities. The experimental survey is a fruitful instrument for testing different dimensions of how organisations function as complex systems – such as gender bias in the hiring process or within horizontal mobility (between different departments/areas) or vertical mobility (in the organisational hierarchy). However, the difficulty in controlling all contributing factors that can

weigh on respondents' response patterns also recommends the triangulation with other methodologies (Dutz et al., 2023).

Acknowledgements

The authors acknowledge the financial support granted by Fundação para a Ciência e Tecnologia (Ref. 2022.08793.PTDC).

References

- Barron, K., Dittmann, R., Gehrig, S. and Schweighofer-Kodritsch, S. (2022) "Explicit and Implicit Belief-Based Gender Discrimination: A Hiring Experiment", *SSRN Electronic Journal*.
- Bendick, M., and Nunes, A.P. (2012) "Developing the Research Basis for Controlling Bias in Hiring", *Journal of Social Issues*, Vol. 68, No. 2, pp 238–262.
- Birkelund, G.E., Lancee, B., Larsen, E.N., Polavieja, J.G., Radl, J. and Yemane, R. (2022) "Gender Discrimination in Hiring: Evidence from a Cross-National Harmonized Field Experiment", *European Sociological Review*, Vol. 38, No. 3, pp 337–354.
- Casaca, S., G., M.J., Marques, S. & Paço, N. (2020) "WoB metro 2018-2019. Factsheet of the Project Women on Boards - an integrative approach". <https://womenonboards.pt/en/wobometr>.
- Carlsson, M., Finseraas, H., Midtbøen, A.H. and Rafnsdóttir, G.L. (2021) "Gender Bias in Academic Recruitment? Evidence from a Survey Experiment in the Nordic Region", *European Sociological Review*, Vol. 37, No. 3, pp 399–410.
- Dezsö, C.L., Ross, D.G. (2012) "Does female representation in top management improve firm performance? A panel data investigation", *Strategic Management Journal*, Vol 33, pp 1072-1089.
- Dutz, R., Hubner-Benz, S., Emmerling, F., and Peus, C. (2023) "Sure you are ready? Gendered arguments in recruitment for high-status positions in male-dominated fields", *Frontiers in Psychology*, Vol 13, pp 1-19.
- Eagly, A. H., Nater, C., Miller, D. I., Kaufmann, M., & Sczesny, S. (2019) "Gender Stereotypes Have Changed: A Cross-Temporal Meta-Analysis of U.S. Public Opinion Polls From 1946 to 2018", *American psychologist*, Vol. 76, No. 3, pp 301-305.
- Erlandsson, A., Bygren, M. and Gähler, M. (2023) "Is there a rating bias of job candidates based on gender and parenthood? A laboratory experiment on hiring for an accounting job", *Acta Sociologica*.
- Friedmann, E. and Efrat-Treister, D. (2023) "Gender Bias in Stem Hiring: Implicit In-Group Gender Favoritism Among Men Managers", *Gender and Society*, Vol. 37, No. 1, pp. 32–64.
- Gastaldi, L. & Corso, M. (2014). "Academics as effective orchestrators of interorganizational change and development: the experience of Italian observatories", *Research in Organizational Change and Development*, Vol. 21, pp. 59-119.
- Hoover, A.E., Hack, T., Garcia, A.L., Goodfriend, W. and Habashi, M.M. (2018) 'Powerless Men and Agentic Women: Gender Bias in Hiring Decisions', *Sex Roles*, Vol. 80, No. 11–12, pp 667–680.
- Kessler, J.B., Low, C. and Sullivan, C. (2019) "Incentivized resume rating: Eliciting employer preferences without deception", *American Economic Review*, Vol. 109, No. 11, pp 3713–3744.
- Malik, M. & Makhdoom, D. (2016). Does corporate governance beget firm performance in Fortune Global 500 companies?, *Corporate Governance* 16(4), pp. 747-764.
- MacNell, L., Driscoll, A., Hunt, A. N. (2015) "What's in a Name: Exposing Gender Bias in Student Ratings of Teaching", *Innovative Higher Education*, Vol. 40, pp 291-303.
- Munsch, C.L. (2016) "Flexible Work, Flexible Penalties: The Effect of Gender, Childcare, and Type of Request on the Flexibility Bias", *Social Forces*, Vol. 94, No. 4, pp. 1567–1591.
- Ramos, A.M. and Félix, B. (2019) "Efeitos do Gênero sobre a Decisão de Contratação e Promoção de Líderes", *Revista Ibero-Americana De Estratégia*, Vol. 18, No. 1, pp. 71–89.
- Seierstad, C. (2016) "Beyond the business case: The need for both utility and justice rationales for increasing the share of women on boards", *Corporate Governance. An International Review*, Vol. 24, No. 4, pp. 390-405.
- Terjesen, S., Sealy, R., & Singh, V. (2009). Women directors on corporate boards: A review and research agenda. *Corporate Governance: an International Review*, Vol. 17, No. 3, pp. 320-337.
- Terjesen, S., and Sealy, R. (2016) "Board gender quotas: Exploring ethical tensions from a multi-theoretical perspective", *Business Ethics Quarterly*, Vol. 26, No. 1, pp. 23-65.
- Tremmel, M., and Wahl, I. (2023) "Gender stereotypes in leadership: Analyzing the content and evaluation of stereotypes about typical, male, and female leaders", *Frontiers in Psychology*, Vol. 14.