

Vignette study for the analysis of labor recruitment: A critical perspective

Juan-Ignacio Martínez-Pastor
Irina Fernández-Lozano

UNED. Sociology Department II
jimartinez@poli.uned.es; irina.fernandez@poli.uned.es



Received: 11-01-2022
Accepted: 13-06-2022
Published: 20-10-2022

Recommended Citation: Martínez-Pastor, Juan-Ignacio and Fernández-Lozano, Irina (2022). "Vignette study for the analysis of labor recruitment: A critical perspective". *Papers*, 107 (4), e3075. <<https://doi.org/10.5565/rev/papers.3075>>

Abstract

A vignette study is, in essence, an *ad hoc* survey designed to evaluate fictitious scenarios on the part of respondents, so that opinions or attitudes can be inferred from the evaluation that are not usually revealed in traditional surveys. The main advantages of vignettes are twofold: firstly, by virtue of the survey design, existing correlations between variables that always occur in real populations are eliminated and the effects of these variables can be better disentangled; and secondly, that respondents respond indirectly to the research questions, so this technique tries to eliminate social desirability biases. After reviewing the basic characteristics of the technique, the article considers the usefulness of these studies from a critical viewpoint in relation to the following aspects: social desirability biases; the fact of posing hypothetical situations and not analyzing real behavior and, above all, their external validity when applied to very specific groups (or the extent to which research results can be generalized).

Keywords: vignette; vignette studies; discrimination; gender; internal validity; external validity; factorial surveys; factorial study

Resumen. *El estudio de viñeta para el análisis de las contrataciones laborales: una perspectiva crítica*

Un estudio de viñeta es, en esencia, una encuesta diseñada *ad hoc* consistente en evaluar escenarios ficticios por parte de las personas encuestadas, de tal manera que de dicha evaluación se puedan inferir opiniones o actitudes que en encuestas tradicionales no suelen salir a la luz. Sus ventajas principales son dos: en primer lugar, que, en virtud del diseño de la encuesta, se eliminan las correlaciones existentes entre variables que siempre se dan en poblaciones reales, lo que permite desentrañar mejor los efectos de esas variables; y, en segundo lugar, que las personas encuestadas responden indirectamente a las preguntas de investigación, por lo que esta técnica trata de eliminar sesgos de deseabilidad social. Tras revisar las características básicas de la técnica, el artículo plantea su utilidad desde un punto de vista crítico en relación con los siguientes aspectos: con los sesgos de deseabilidad social, con el hecho de plantear situaciones hipotéticas y no analizar comportamientos reales y, sobre todo, con su validez externa cuando se aplica a colectivos muy concretos (o hasta qué punto se pueden generalizar los resultados de este tipo de estudios).

Palabras clave: viñeta; estudios de viñeta; discriminación; género; validez interna; validez externa; encuestas factoriales; estudio factorial

Abstract

- | | |
|---|------------------|
| i. Introductio | 4. The drawbacks |
| 2. What is a vignette study? | 5. Conclusions |
| 3. What are the advantages of vignette studies? | Acknowledgments |
| | Funding |

1. Introduction

Experiments enjoy a long tradition in some areas of the social sciences, such as psychology or market research. Although this is not the case in sociology, the use of experimental designs has become increasingly common in our discipline in recent years. This is due to the degree of maturity reached by the discipline; a maturity that has led to a critical review of the methodology applied so far and a need to go beyond the limitations of the most common designs, methods and techniques.

This critical review has resulted, on the one hand, in the blurring of the boundary between qualitative and quantitative research in terms of research design and inferential logic (King et al., 2005) and, on the other, in the desire of researchers to go beyond the discovery of correlations and a superficial description of the facts. Indeed, sociological researchers, tired of correlations and devoid of explanations, are focusing on how and why social events happen, on establishing their causal mechanisms. This is where experimental or quasi-experimental designs become relevant (Wallander, 2009).

This article is based on the experience and critical reflections of the authors after having applied one of the most widely employed quasi-experimental techniques in the last decade: vignette studies.¹ The article is divided into four

1. The research was published in a first quartile JCR sociology journal (Fernández-Lozano et al., 2020).

parts. The first offers a brief description of vignette studies and the second presents the main advantages of this technique. The third section, which is more extensive, is devoted to a critical analysis of the drawbacks to the technique, particularly the problem of external validity due to the samples that are usually used. Lastly, we include a brief section with conclusions. It is necessary to emphasize that much of the criticism regarding the external validity of vignette studies focuses on the analysis of labor recruitment, so it is not a criticism that can be generalized to other thematic areas in which this technique has been applied, such as attitudes toward basic income (Laenen et al., 2022), redistributive policies (Gniza et al., 2022) or prostitution (Escot et al., 2021), to cite just three examples.

2. What is a vignette study?

In essence, a vignette study is a survey that presents fictitious scenarios to respondents for evaluation to glean information about their opinions or attitudes. Thus, vignettes differ from traditional surveys mainly in their design phase, but not so much in the analysis phase (although, as we will see below, the design of the vignette itself enables a multilevel analysis, which does not necessarily occur in other types of surveys). Often, such fictitious scenarios or situations present the particular feature that they have been constructed using all possible combinations of a series of dimensions² (variables) whose impact on opinions, attitudes or even potential behaviors is to be analyzed (Auspurg and Hinz, 2015). These dimensions are variables that researchers consider socially sensitive in relation to the topic of study and would therefore be very problematic to ask about openly. Vignette studies have two main advantages. First, through an orthogonal construction of the fictitious scenarios (often people), the existing correlations between variables in real populations are eliminated, thus making it possible to better isolate their effects. Second, to eliminate social desirability biases, respondents ideally respond indirectly to the research questions. In other words, as in all other experimental or quasi-experimental techniques, respondents are subjected to a *treatment* of which they are unaware. Some examples of these indirect questions are provided in Section 2.1. In what follows, the design phases of vignette studies are described in detail.

2.1. Choice of theme, hypotheses and dimensions

As in all social research, the choice of theme and hypotheses must precede the choice of technique. Therefore, prior to choosing to perform a vignette study, it should be compared to other alternatives to determine if this type of

2. Strictly speaking, this case would refer to factorial surveys, while the term *vignette* refers to the presentation of the fictitious situation. However, the terms *vignette* and *factorial surveys* are often used interchangeably in the literature. In this article we focus on vignette studies based on factorial surveys.

study best suits the research question. The suitability of a vignette design is, at least in theory, simple to assess. Sensitive questions one would not usually answer openly are conducive to this type of study (i.e., “Would you hire a pregnant woman in your company?”, “Would you mind having a person from group X as your neighbor?”; “Who would you prefer to receive social assistance from the municipality?”). Obviously, very few people would dare to say that they *discriminate* against a certain group in a survey. On the other hand, if there are no problems of social desirability, a vignette (despite the complexity involved in its design) would probably not provide any additional value compared to a direct question about the issue. Moreover, vignette surveys are not valid for inquiring into specific past events (even if they are subject to social desirability), but rather into attitudes, opinions and even potential behaviors, since the vignettes represent fictitious scenarios created specifically for the research.

As we will see below, one of the typical uses of vignette studies is to evaluate fictitious job applications, which is the focus of much of this article. For this reason, these fictitious scenarios often take the form of somewhat complex resumes or profiles. Di Stasio analyzed which resumes were considered most presentable from the perspective of Italian employers (Di Stasio, 2014a). Karpinska et al. (2013) assessed the characteristics older workers should have to increase their likelihood of remaining with the company in the Danish context. Gender biases or those particularly associated with motherhood in hiring or job promotion have been the subject of several investigations in the United States (Benard and Correll, 2010; Correll et al., 2007; Cuddy et al., 2004). However, the object of study of a vignette is not limited to the assessment of a resume by potential or actual employers. Any topic lends itself to the use of a vignette if what is of interest is the identification of prejudice, stereotypes or discrimination. In this line, Kootstra (2016) analyzed the extent to which the British and Dutch considered individuals of different ethnic and migrant backgrounds as “deserving” of public benefits. Jacobs and Gerson (2016) studied the opinions of Americans in relation to mothers in paid work. The choice of dimensions or variables to be studied bears a close relationship to the formulation of the hypotheses, since fictional scenarios referring to individuals, ethnicity, gender, sexual orientation, parental status or age are typical dimensions in vignette studies.

2.2. Designing vignettes

The design of fictitious situations or vignettes basically involves three phases: two of a more technical nature and a final, more creative phase. First, a universe of vignettes is created from all possible combinations of categories (or “levels”) within each dimension. Thus, the larger the number of dimensions to be tested and the larger the number of categories within these dimensions, the larger the size of the universe of vignettes and hence the larger the sample size necessary to analyze the effect of these dimensions while avoiding correlations

that occur in real life.³ In short, a fictitious population is created where there is no correlation between, for example, immigrant status and income level, or gender and level of responsibility within a company, since all the main independent variables respond to an orthogonal or Cartesian design. Thus, in our fictitious universe, women and men would have the same probability of being managers or subordinate employees, so that the effect of both variables on our dependent variable can be disentangled.

The second design phase consists of assigning the vignettes to different questionnaire models, which involves deciding how many vignettes each respondent will evaluate.⁴ Ideally, if each person evaluates one candidate, the fatigue effect is avoided and the dimensions to be tested are better concealed, since they are not repeated. However, this type of design is more costly, since it considerably increases the sample size required to achieve statistical efficiency (perhaps not by chance, the number of vignettes evaluated per respondent is generally lower in North American studies than in European ones). For this reason, most studies present between two and six vignettes. From the complete universe of vignettes (simply the Cartesian combination of all the categories of all the variables, as shown in Table 1), it may be necessary to extract a sample of profiles to be included in the questionnaires. Most studies use a random sample, although some authors have suggested the convenience of using a quota sample (Dülmer, 2007). Statistical analysis programs can be used to obtain and distribute the sample in questionnaire models, in such a way as to minimize the correlation between variables and maximize statistical efficiency.⁵

Finally, the questionnaire is designed, including a cover story (a narrative that serves as a “pretext”) to conceal the ultimate objective of the research and not put the respondents “on alert”. Any additional element introduced in the narrative must be neutral in relation to the research objective and can be either common to all the vignettes or controlled as an additional dimension by the researcher.⁶ Great care must be taken when designing a vignette, because any element of the narrative or the way the vignette is presented can potentially introduce an uncontrolled variable that correlates with the dependent variable.⁷ Beyond that, there are multiple ways to present the situation to be assessed.⁸ Often, dimensions are presented indirectly. For

3. As we will see below, a sufficient sample size is also necessary to ensure the external validity of the research.
4. “Dummy” vignettes (common to all questionnaires and which will not be part of the central analysis) are sometimes included to avoid the order effect or other biases (see Castilla and Bernard, 2010, who added a third vignette to conceal that gender was one of the dimensions being analyzed).
5. Researchers usually use the SAS ‘%mktblock’ macro.
6. Control dimensions not directly related to the research hypotheses are often introduced to add complexity to the vignettes or to rule out spurious correlations.
7. For example, the order in which the vignettes are presented (see the review by Auspurg and Jäckle, 2015).
8. Always with the double purpose, which is sometimes difficult to reconcile, of concealing the treatment to avoid problems of social desirability and respondent fatigue.

Table 1. Examples of vignette presentations in promotion, recruitment and personnel evaluation studies.

	Vignettes using questionnaires (real + “fillers”)	Total no. of profiles*	Ratio respondents/ profiles	Presentation of profiles
Wolf and Velden (2001)	20	$2 \times 3 \times 6 \times 2 \times 2 \times 3 \times 2 \times 4 \times 6 \times 2 \times 2 \times 2 = 331.776$ (selected a sample universe of 138,240 realistic profiles)	<1	Resumes and cover letters that varied by <u>gender</u> , <u>age</u> and <u>type of education</u> (2 dimensions); <u>academic outcomes</u> (2 dimensions); <u>work experience</u> , <u>extra-curricular activities</u> (2 dimensions); <u>4 personal characteristics and appearance</u> (using a pre-tested photograph).
Cuddy, Fiske and Glick (2004)	3 (1 + 2)	$2 \times 2 = 4$	30	“ <u>Kate (Dan)</u> is a 32-year-old associate consultant who graduated with an MBA. She’s (He’s) been working in her (his) current field for six years. When working with a client, her (his) duties include identifying problems, planning and conducting interviews and analyses, synthesizing conclusions into recommendations, and helping to implement change in her (his) client’s organizations. Her (his) hobbies include swimming and tennis. (<u>Kate and her husband [Dan and his wife] recently had their first baby.</u>) She (He) lives in central New Jersey, commuting to work two days a week and telecommuting three days a week”
Castilla and Bernard (2010)	3 (2 + 1)	$2 \times 2 = 4$	29/57/25 (3 samples)	Performance and staff development evaluation form that included the <u>gender</u> of the person and a description of the company as <u>meritocratic or non-meritocratic</u> . It also included a quantitative and qualitative assessment of employees’ performance, which was equivalent in all profiles.
Rudman and Mescher (2013)	1	$2 \times 4 \times 2 = 16$	23	Transcript of an interview between an employee and a human resource manager that varied by the <u>type of leave requested</u> (<u>childcare/parental/none</u>), <u>target race</u> (<u>Black/White</u>) and <u>gender</u> .
Di Stasio (2014a)	18	$2 \times 2 \times 2 \times 3 \times 3 \times 2 \times 3 \times 2 = 864$	<1	Resumes that varied in 8 dimensions (<u>gender</u> , <u>previous work experience</u> , <u>completion of company internship</u> , <u>educational level</u> , <u>field of study</u> , <u>academic record</u> , <u>duration of studies</u> , <u>extra-curricular activities</u>).

Note: The dimensions are underlined.

* from the Cartesian combination of all categories of all dimensions.

Source: Own elaboration.

example, ethnicity or gender can be presented through the person's name, as in Correll et al. (2007) or Fernández-Lozano et al. (2020). The questionnaire to evaluate the vignette must also include one or more closed-ended questions, which are the dependent variables.⁹ These questions often serve as an indirect way to approach the object of study similar to how we approach sensitive issues in our daily lives (e.g., if we ask "How's it going with Pedro?" instead of "What do you think about the way he talks to us?"). Sometimes there is more than one dependent variable, such as when simulating the stages of a real selection process (Wolf and Velden, 2001) or more often because some of the variables are considered mediating variables that elucidate the mental mechanisms behind a particular attitude (Cuddy et al., 2004; Fernández-Lozano et al., 2020).

2.3. Analysis of the results

Once the questionnaire has been administered,¹⁰ a database is obtained where the vignettes are the final units of analysis (often the fictitious candidates, as we have seen) grouped into respondents. This multilevel structure allows us to distinguish the different levels of variance to take into account (between respondents and for the same respondent). As will be seen below, it is important to have a sufficiently large sample of respondents to ensure the external validity of the experiment.¹¹

3. What are the advantages of vignette studies?

Like other experimental techniques, the most important advantage of vignette studies is that the conditions the participants are exposed to in the experiment can be manipulated to test the hypotheses. In addition, because these studies are designed *ad hoc*, what is intended to be measured is in fact measured, without the use of approximations or remediation. These advantages concern the construct validity and internal validity of a study. Construct validity refers "to the correspondence between a theory and a research design that sets out to test that theory", while internal validity refers "to the truth of a proposition with respect to the chosen sample" (Gerring, 2014: 457).

The high internal and construct validity enable establishing causal relationships with much greater certainty than in observational studies. Thus, unlike observational studies, which are more common in the social sciences, in vignette studies the researcher manipulates the set of factors (equivalent to variables in quantitative analyses) and levels (equivalent to the categories of

9. In addition to variables to classify the respondents or others.

10. Vignettes are administered like any other type of survey, although we have found no evidence of vignettes being administered by telephone.

11. For a more in-depth discussion of multilevel analyses of vignette studies, see the methodological appendix in Fernández-Lozano et al. (2020).

Table 2. Main advantages and drawbacks of vignette studies

Advantages	Drawbacks
High internal validity. They serve to unveil causal relationships between social facts.	Hypothetical situations and fictitious candidates are analyzed.
High construct validity.	It is much easier to eliminate all social desirability biases (by “concealing” the treatment) than in a normal survey, although it is not fully guaranteed that this will be achieved.
Vignette studies avoid the problem of candidate self-selection or self-censoring for the subject under analysis (e.g., recruitment), which occurs in the real world. With vignettes, it is possible to know the extent to which self-censored characteristics carry weight in job hiring or promotion processes.	As in other <i>ad hoc</i> surveys, it is very costly to achieve a sufficiently representative sample to ensure external validity
Vignette studies raise fewer ethical issues than other experimental designs, such as auditing.	A real situation is not being evaluated.

Source: Own elaboration.

these variables) presented to the respondent in such a way that correlations between factors are minimized to a value close to zero. As noted, this allows separating the effects of variables correlated in the real world as the problem of multicollinearity is avoided and a causal interpretation of the relationships between variables can be made (Di Stasio, 2014b: chapter 4).

Vignette studies also present advantages over *riskier* experimental studies that simulate real situations (completely concealing the conduct of research), such as audit studies,¹² which have more ethical implications. Participants in vignette studies are always fully aware that they are participating in social research, although the true purpose of the research is not usually revealed to them. Moreover, the cost of vignette studies is the same as that of *ad hoc* surveys because they are, in essence, surveys.

Another advantage of vignette analyses is that, for some topics in sociology where this method is used, such as job recruitment or promotion analyses, the technique avoids the problems of candidate self-selection. By randomly combining the levels of the vignettes, any candidate can apply for a position. In the real world, however, many candidates do not apply for a job offer because they believe they have no chance of being chosen for the position. The vignette technique allows testing to what extent they may or may not be chosen (Di Stasio, 2014b: chapter 4).

12. Audit studies, also called “correspondence studies”, involve simulating job applications in a real-life context (usually by sending fictitious resumes to employers) and therefore cannot be considered survey-based experiments. See, for example, Booth and Leigh (2010), Cortina et al. (2021) and Gaddis (2015).

4. The drawbacks

4.1. *Hypothetical situations and social desirability*

The advantages of this technique do not hide its drawbacks. Although we will focus on disadvantages related to the samples used in these studies, an issue that is usually treated very superficially in the literature, it is necessary to mention other important drawbacks. The first is that the respondents evaluate hypothetical situations, not real ones. For example, in vignette studies on the hiring or promotion of personnel, fictitious workers are presented, as is the hiring or promotion itself. This is a disadvantage compared to other experimental techniques such as auditing, where job candidates are also fictitious, but the respondent is unaware of this, and the job offers are real. As the use of vignettes is a relatively recent technique in sociology, few studies have analyzed the effects of presenting hypothetical rather than real situations. In this regard, some studies have found no significant difference between evaluating one type of situation or the other, such as Hainmueller et al. (2015) in an experiment on attitudes towards immigration in Switzerland and Gutfleisch et al. (2021) in an experiment on labor recruitment in Luxembourg.

In a similar line, another major drawback concerns the possible social desirability biases involved in vignette studies (McDonald, 2019). It has been argued that this technique is more appropriate than surveys precisely to avoid such biases, as they do not inquire into sensitive topics and present a hypothetical situation in which the true purpose of the study is often masked (Sterkens et al., 2020). For this reason, vignette studies are often used to analyze attitudes. However, it could happen that those who respond to vignettes by deciding to hire or promote a woman rather than a man or a mother with young children rather than someone who has no children are permeable to political correctness. Indeed, although the purpose of the study is not made explicit and is not as readily apparent as in a normal survey, there is no guarantee that the key differences between the vignettes will go unnoticed by human resource (HR) experts and they will respond due to social desirability, contrary to what they would do in real decision-making scenarios. This same example can be extended to other issues related to attitude analysis.

In a study on gender inequality in hiring, Kübler et al. (2018) found a solution to this potential bias. The solution was to have each HR expert evaluate a group of vignettes with candidates of the same gender, rather than presenting candidates of different genders to the same expert. In other highly imaginative research, Eifler and Petzold (2019) attempted to measure whether individuals' actual behavior differs from their vignette responses. To do so, the authors examined the behavior of drivers when blocked by a car at a stop light when the signal turns green. They then applied a vignette study to the same population. The analysis found that, in general, when responding to the vignettes, people said they would behave in a less aggressive manner than how they actually behaved in a real-life situation. In short, there is still insufficient

empirical evidence to determine the effects of presenting hypothetical rather than real-life situations.

4.2. *The problem of external validity: to what extent can the results of a vignette study be generalized?*

In addition to the problems related to social desirability and the fact that vignettes are made-up situations and not real ones, there is another problem that goes largely unnoticed and is not usually questioned in vignette studies applied to very specific groups, such as those responsible for recruitment and promotion: the external validity of these experiments derived from the sampling characteristics. To what extent can the findings of a vignette study on labor recruitment be generalized beyond the sample used? To what extent can the conclusions be applied across populations and situations?¹³

Some researchers have pointed out that the features of vignette studies that may limit their external validity are the choice of factors and vignette levels, their mode of presentation and what is asked (Karren and Barringer, 2002; Steiner et al., 2016). In fact, these are issues that can arise in any survey-based study. It is well known that the way the questions are formulated, what they ask about, the order in which they are asked, and the response anchors used in closed-ended questions can influence respondents' answers.

In what follows, we discuss the difficulties involved in using vignette studies when they are based on *ad hoc* surveys of very specific groups. These difficulties refer to the collection of data by means of a rigorous sampling procedure as occurs in surveys habitually used in academia. In our opinion, this greatly conditions the external validity of this technique for some analyses, such as those that focus on the recruitment or promotion of personnel. Our argument is based on the following premises, which are well known to experts in sociological surveys and should always be considered when analyzing a survey. The first is that a survey does not tell us exactly what the population under study thinks, but gives us an *approximate* idea, with a margin of error. Moreover, and this is the main point, a survey gives us an approximate idea provided it fulfills an essential condition: it must be *representative*. To this end, if the population under study is segmented into groups, the sample must have the same percentage of each group that we believe the population has (Martínez Pastor, 2019).

13. The methodologist John Gerring (2014: 106/85) argues that the distinction between internal and external validity is crucial in any type of analysis, although it is not always so clear: "As an example, consider a hypothetical study of a school district in the state of New York that rests on a sample of students drawn from that district, but purports to elucidate features of all schools within the state. This presents three potential levels of validity: (1) the sample of students; (2) the school district; and (3) schools throughout the state (across multiple districts). Internal validity may refer to (1) or (2), while external validity may refer to (2) or (3)." Gerring notes that the claim to validity seems less likely as the circles expand, but none is entirely plausible, and from this perspective, there is no clear demarcation between internal and external validity.

Although pure chance through simple random sampling is the best guarantee that a sample will be representative, there are other types of sampling that are appropriate when a complete list of the study population is not available.¹⁴

Despite the advances made in vignette studies, there are certain drawbacks to this technique that affect the generalization of results. These drawbacks are not due to the technique itself—since what defines the vignette, as we have seen, is how the survey is designed and not how the sample is selected—but to the practical limitations of vignettes when applied to truly representative samples aimed at very specific groups that are difficult to locate and convince to collaborate in the survey. Although this technique is applied to many analyses of the general population, it is also targeted at very small groups such as people in charge of personnel recruitment, the subject of this article. In this case, the following aspects should be noted regarding the external validity of this type of vignette study:

- 1) Many of the samples are not truly random.
- 2) In the case of random sampling, there is no certainty as to whether the list from which the sample is drawn is exhaustive (i.e., whether it contains the entire group to be analyzed).
- 3) Given the specificity of the groups that are usually analyzed, it is very difficult (and almost impossible) to stratify the sample or establish quotas based on the individual characteristics of the respondents.
- 4) Small-*N* problem. Most vignette studies targeting very specific groups are based on very few participants, largely because of limited resources.
- 5) The response rate is usually very low.

To support these ideas, we rely on the research compiled by McDonald (2019), who searched the Google Scholar and JSTOR academic databases for sociological studies on labor recruitment using the vignette technique with samples of employers or HR professionals. The list contains 16 studies, to which we have added the study by Fernández-Lozano et al. (2020). The following table, compiled by us based on McDonald's list, includes these investigations and describes the characteristics of the samples broken down into five sections: a) who the sample is targeted at; b) the type of sampling used; c) the sample size; d) the observations analyzed (or the total number of vignettes evaluated by the participants); and e) the response rate.¹⁵ The fact that McDonald chose to focus on recruitment research is not by chance, since experts on this topic have often used vignette studies in recent years.

14. For a discussion on the importance of sample representativeness, see Martínez Pastor (2019: chapter 7).
15. McDonald (2019: appendix 2) includes the sample size in a separate table. McDonald's figures do not match the figures in our table, since we have counted the cases that were actually included in the analyses and not the participants who responded to the survey but did not pass the researchers' filters and were not included in the analyses that were finally published.

Table 3. Characteristics of samples in vignette studies

	Targeted sample	Type of sampling	Sample size	Observations (total vignettes evaluated)	Response rate (final sample used)
Beek et al. (1997)	General managers or personnel with recruiting experience (except for 4% with no recruiting experience). The Netherlands.	Representative sample of Dutch companies from a list. Large companies are overrepresented. The authors do not indicate if the sampling was random or not.	312	5616	41%. 758 were contacted and 475 responded. Final sample used: 312.
Wolf and Velden (2001)	Managers of personnel departments that hired graduates from the Faculty of Social Sciences at Utrecht University.	It is inferred that stratified random sampling was performed using a list of entrepreneurs compiled by the university. Stratification is done by company size, region and for-profit and non-profit organizations. It is not clear how the list was compiled.	27	620	65%
Biesma et al. (2007)	Employers of recent graduates from a public health organization in the Netherlands.	It is inferred that random sampling was performed using a list of potential employers of public health graduates. It is not clear where the list came from.	168	315	18%. 1040 employers were contacted by mail; 185 responded and information from 168 was finally used.
Henkens et al. (2009)	2 samples: 1) Managers from 4 large Dutch multinational companies. 2) Business school students.	Twenty-six managers from four large Dutch multinational companies and 25 business school students were selected. The selection process is not indicated.	26 managers and 25 students	611 and 610 vignettes in each study	—
Karpinska et al. (2011)	2 samples: 1) Managers attending a focus group on the return to employment of older workers and others captured by the snowball method. 2) Business students at Utrecht University.	1) Manager sample: snowball sampling from focus group attendees. 2) Student survey. Type of sampling not specified.	20 managers and 17 students	240 and 204 for each study	—
Karpinska et al. (2013)	Supervisors of the Longitudinal Internet Studies for the Social Sciences (LISS panel) of Tilburg University.	Managers of the survey were contacted.	238	1190	Exact response rate not known. The authors say they contacted 700 supervisors of whom 515 responded, but then did another filter to differentiate the managers. A total of 238 managers participated, but the exact data to calculate the response rate is not known.
Di Stasio (2014a)	Recruiters from HR offices in a specific sector of activity: NACE code = 62 (computer programming, consultancy and related activities). Italy.	Random sampling. 148 companies in the sector registered in the Chamber of Commerce were contacted.	59	1062	40% response rate. The study includes a table of the sample representativeness for the designed sample and the entire sector by company size and geographic location (north, center and south).
Mulders et al. (2014)	Management team of the LISS panel of Tilburg University.	The authors contacted the LISS management team.	443	2215	The exact response rate is not known. The authors state they contacted 864 individuals and 698 (81%) responded, but then filtered the participants based on several questions to make sure the managers had been involved in recruiting processes during the past 10 years. Data from 443 individuals were finally used.

Table 3. Characteristics of samples in vignette studies (*continuation*)

	Targeted sample	Type of sampling	Sample size	Observations (total vignettes evaluated)	Response rate (final sample used)
Mulders et al. (2014)	Management team of the LISS panel of Tilburg University.	The authors contacted the LISS management team.	443	2215	The exact response rate is not known. The authors state they contacted 864 individuals and 698 (81%) responded, but then filtered the participants based on several questions to make sure the managers had been involved in recruiting processes during the past 10 years. Data from 443 individuals were finally used.
Humburg and Velden (2014)	People involved in recruitment processes in the last 5 years in one of 7 sectors of activity chosen by the researchers. Study of 9 European countries.	The authors invited all individuals who met the criteria to participate based on a panel of consumers registered with an international market research company.	903	9030	20% response rate, although it is not a response rate in the strict sense, since when 100 responses were reached in a country, they stopped administering the survey.
Di Stasio and Gërzhani (2015)	HR professionals in the ICT (information, communication and technology) sector. England.	Random sampling from companies affiliated with two associations in the ICT sector (one of small and medium-sized companies and the other called the “technology industry”).	22	220	20% response rate. 172 companies were contacted and 34 participated, but throughout the article and in the tables the authors note that 22 companies participated in the survey. They provide a table comparing the size of the companies they contacted and the size of the companies that responded. They only have information to provide this comparison.
Karpinska et al. (2015)	Management team of the LISS panel of Tilburg University.	The authors contacted the LISS management team.	153	765	The exact response rate is not known. The authors state they contacted 700 supervisors and 515 (74%) responded. After applying filters to ensure that they were managers and not supervisors, 238 remained, of which 153 participated.
Dame- lang and Abraham (2016)	CEOs or entrepreneurs of the Cluster-Oriented Regional Information System (CORIS) database. Germany.	The authors invited all the companies in the CORIS database.	187	997	11% response rate. 187 out of the 1474 companies that were contacted responded (12%). Data from 168 people were finally used.
Di Stasio and Werfhorst (2016)	Recruiters and HR professionals at IT companies in the Netherlands and England.	England: Random sampling from companies affiliated to two associations in the sector (one of small and medium-sized companies and the other called the “technology industry”). The Netherlands: Random sampling. 148 companies belonging to the sector and registered with the Chamber of Commerce were contacted.	38 in the Netherlands and 34 in England	684 and 612, respectively	The authors contacted a total of 172 companies and 72 responded. They state that there is supplementary material in the journal to analyze the sample representativeness, but we have not been able to access it through the link they provide.
Liechti et al. (2017)	Hotel managers. Switzerland.	All members of the main Swiss hotel business association were invited.	243 and 232 depending on the analysis	967 and 931 depending on the analysis	12% response rate. 1,982 managers were invited and 238 participated, although depending on the analysis, the tables indicate that 232 to 243 participated.

Table 3. Characteristics of samples in vignette studies (*continuation*)

	Targeted sample	Type of sampling	Sample size	Observations (total vignettes evaluated)	Response rate (final sample used)
Oesch et al. (2017)	HR managers. Switzerland.	4687 HR managers from the largest human resource management association in Switzerland were invited to participate.	385	1644	8% response rate. 4687 HR managers were invited and 714 responded (15%). They finally used information from 385 recruiters after filtering for non-response and for the fact that the recruiters were currently employed.
Shi et al. (2018)	Recruiters of real job vacancies advertised on web portals in German-speaking Switzerland. Occupations related to mechanics, banking and insurance, gastronomy, nursing and ICT were selected to include low, medium and high-skilled jobs.	It is inferred that all the recruitment officers for the actual job vacancies specified in the left column were invited.	638 and 634, depending on the analysis, although in a note they state that 580 responded.	6338 to 6298	27% response rate. 2118 recruitment officers were invited to participate. 739 began the survey and 580 completed it.
Fernández-Lozano et al. (2020)	Supervisors of companies with at least 30 employees in three sectors according to the degree of feminization of the workforce: programming, consulting and other computer-related activities; retail trade; and advertising and market research.	Random sampling from a database of companies purchased from a commercial firm.	71	426	14% response rate

Source: Own elaboration based on the list of publications compiled by McDonald (2019) and Fernández-Lozano et al. (2020).

Several conclusions can be drawn from the analysis of the characteristics of the samples used in these 17 studies on recruitment:

- 1) The Netherlands is overrepresented. Eight of the studies were carried out exclusively with Dutch data.
- 2) When writing the articles, the authors dedicate more time describing the design of the vignettes in detail (which is undoubtedly important) than on explaining the sample characteristics. In some cases, the type of sampling must be inferred. In the case of random sampling, it is not usually clear what type of list is used to select the companies and, above all, whether they are exhaustive lists of the group studied.
- 3) Ten studies used lists of companies or professionals belonging to associations. Of these, six used random sampling from these lists and four others attempted to contact all the companies on the lists so that a manager or HR professional could participate in the experiment.
- 4) Three other studies used surveys already conducted in the general population to contact all managers or HR professionals who participated in those surveys.

- 5) The rest of the studies were based on
 - a. A sample obtained from a focus group and another sample of students using the snowball method.
 - b. A consumer panel.
 - c. Real job vacancies advertised on web portals.
 - d. The information provided in another study does not allow us to determine how the sample was obtained.
- 6) The largest sample size is 903 individuals, while the smallest one is just 22. The mean sample size for the 17 studies is 235 individuals.
- 7) In some studies, it is not possible to determine the exact response rate. In those where it is clear, the response rates range from 8% to 65%. The average response rate (calculated on the cases finally used in the analyses) is 27%.

Of the above points, the following problems stand out. In random sampling, it is not usually known whether the list of companies contacted is exhaustive or not. The same is true of the associations in which managers or HR professionals are members. If the lists from which the samples are drawn are not exhaustive, the question arises as to whether there is any bias in the characteristics of the companies or recruiters on the lists. If this bias affects hiring preferences, the analysis will surely be biased.

Added to this is the impossible task of stratifying the sample according to the characteristics of managers, HR professionals or those in charge of selecting or hiring personnel. This is because, unlike general populations, whose basic characteristics can be determined from censuses or population registers, there are usually no exhaustive records with the characteristics of all individuals in a given occupation.¹⁶ Some of the studies analyzed stratify the sample by the size of the company, the region or whether the companies are for-profit or not-for-profit, but if biases come into play when hiring, it is plausible that they are not only due to the characteristics of the company, but also the characteristics of those in charge of selecting personnel. If the basic characteristics of a group of professionals are not really known, it is impossible to stratify the sample by individual traits and, in fact, this does not occur.

Another issue that limits the external validity of vignette studies in the field of personnel recruitment is the sample size, which is usually very limited. This is not surprising, since the target groups for these studies are very specific and specialized, such as managers, so they are usually not easy to contact and, if they are, it is not an easy task to get them to agree to participate in experiments of this type. As noted, the average sample size of the 17 studies analyzed was 235 individuals. Once the information has been collected, the researchers examine the data using multilevel analysis and sometimes other

16. In some very large occupations, the characteristics of employed persons may be gleaned from surveys with very large samples, such as the labor force survey of Spain (*Encuesta de Población Activa*; EPA).

regression techniques, with enough cases to carry them out and even to make interactions between variables. It should be noted, as already indicated, that given the multilevel structure of the resulting database, and unlike most survey-based analyses, the cases in this vignette technique are not the individuals who evaluate the vignettes, but the number of vignettes evaluated. In this way, the cases to be analyzed increase significantly, thus allowing analyses with a certain statistical robustness. In the 17 studies cited, the average number of observations was 1938. From the viewpoint of sampling, however, it is not the number of vignettes evaluated that is relevant, but the representativeness of the individuals who have evaluated them. In this regard, seven of the 17 analyses studied included less than 100 participants, and another three included over 100 but did not reach 200. Although the groups analyzed are very specific, these numbers are far below the surveys usually used by academic sociologists to reach reliable conclusions.¹⁷

An additional problem with sampling in vignette studies applied to very specific groups is the low response rate. As academics, we tend to use surveys with probability sampling, which are generally conducted by public institutions due to their high cost and the access these institutions have to population lists on which to design the sample. Once designed, it is very important to achieve a high response rate, since otherwise there may be a self-selection bias among respondents. In Spain, it is compulsory to respond to all surveys administered under the National Statistical Plan, which include all those conducted by the National Statistics Institute (Instituto Nacional de Estadística; INE) and, for sociology, some surveys conducted by the Center for Sociological Research (Centro de Investigaciones Sociológicas; CIS). The importance of obtaining high response rates is such that the institutions conducting these surveys attempt to contact the selected individuals several times before giving up. On occasion, replacements with sociodemographic profiles identical to those of the individuals who cannot be reached are used.

To give an idea, the response rate in Spain for a survey with probability sampling without replacement, such as the European Social Survey, was almost 70% from the third to the eighth wave. In the first two waves, the response rate was 50–60%; in the eighth, the average response rate for the 23 participating countries was 55% (CIS, 2019). As mentioned in the sample characteristics, of the 17 vignette studies analyzed, the response rate in the random samples is much lower and, as can be seen in Table 3, the samples with high response rates have a very small number of cases. Indeed, in the random surveys that achieved the highest rates (above 40%) few individuals were interviewed; the one with the highest response rate interviewed 72.

17. Another issue concerns data misuse. This occurs in surveys involving subsamples of certain groups with larger sampling errors than those reported in the data sheets, since these sheets are based on the total sample size and not on the sample subsets. Although experiments in other fields of research related to medicine, psychology or neuroscience are usually carried out with a very small group of individuals and are held as valid, this is not the case in conventional sociology, which studies general populations.

It is well known that the response rate in non-probability sampling, such as quota sampling, is often lower. This issue is of concern to companies that collect data. The Pew Research Center (PRC), one of the most prestigious institutions in the United States that conducts surveys, took a close look at this problem and found a dramatic decline in telephone survey response rates from 36% to 6% over the period 1997–2018 (Kennedy and Hartig, 2019). The PRC concluded that the biases produced by nonresponse can be minimized through convenience quotas and weightings. In fact, the PRC usually weights their surveys using quite a few variables, such as gender, age, region, municipality size, educational attainment, country of origin and ethnicity.

In general population surveys, it is possible to weight cases by these variables if the characteristics of the population are known. However, as noted, this cannot be done in vignette studies applied to very specific groups, since their characteristics are usually not known. Even when weighting by all these variables, it is known that, at least in the United States, survey respondents are those who participate most in the public life of their community, which can lead to biases with respect to the object of study.

In short, vignette studies carried out for specific groups raise the problem of the possibility of extrapolating the results, even among the members of that group. This is a problem of any survey-based research applied to a very particular group, so this criticism of the external validity of vignette studies is not applicable if the technique is intended for general populations with a usual sampling procedure.

4.3. Possible solutions or resignation

Is there a solution to these problems affecting sample representativeness? One solution, which we already mentioned, would be to stratify the samples or even weight them according to the characteristics of the group participating in the experiment. However, this would require knowing these characteristics, which is not easy.¹⁸ Moreover, if such characteristics were known, they could be compared with the final sample in an exercise of transparency. Di Stasio (2014a) and Di Stasio and Gërkhani (2015) did something similar, although not by comparing the characteristics of the participants in the vignettes but rather the characteristics of the companies where they worked.

Another ideal solution would be to embed the vignettes in a normal survey that is representative of the group to be analyzed. This is not usually easy in very specific groups either, such as managers or HR professionals, due to the lack of

18. For example, in research on the behavior of HR professionals in Spain, one option would be to use the Spanish Labor Force Survey (EPA). However, the National Statistics Institute only includes occupations up to 3 digits in the survey, not 4. This is a drawback, since HR professionals are assigned the code 1212 in the National Classification of Occupations of 2011, the one in force at the date of publication of this article. Thus, in the microdata available for researchers, these professionals are placed in category 121 “Managers of administrative departments.” In the fourth quarter of 2020, there were 319 cases with that code.

representative surveys targeted at these groups. However, as we shall see, there are exceptions. In the absence of such specific surveys, what researchers usually do, as indicated in Table 3, is to take advantage of a traditional survey sample to contact the group to be investigated (Karpinska et al., 2013; Mulders et al., 2014; Karpinska et al., 2015). The problem with using a sample subset is the limited number of cases of a given occupation in the original sample; an obstacle that is exacerbated because the experiment is not embedded in the survey itself. Rather, the subjects in the subsample must respond to the invitation and subsequently accept to participate in the vignette study. Thus, the final sample is usually even smaller, further complicating the small- N problem and possible selection bias.

The analysis by Steiner et al. (2016) provides an example of a vignette study that circumvents the small- N problem by designing a sampling method identical to those more common in sociology (gender and age quota sampling) and applying the vignettes to all respondents, which totaled 980. The study addresses attitudes toward the gender pay gap. The researchers were able to use these quotas with confidence because their study universe was not a very specific group, but the Viennese population aged 18 to 65, whose sex and age structure can be easily known.

The sample used by Kübler et al. (2018) is an exception among vignette studies targeting recruiters. They embedded the vignettes in a wave of an annual panel survey conducted by the Federal Institute for Vocational Education and Training (BIBB) in Germany. The survey is completed annually by company owners or HR managers. As they describe in their research, the sample is drawn from a database of the Federal Employment Agency and contains all companies with at least one employee who pays social security contributions. The researchers did not present the vignettes to all the companies in the sample (3450), but randomly selected 680, with a response rate of 98%, which were the companies that finally participated in the experiment. Thus, there is greater certainty that their findings are valid for the German companies as a whole given the original sample design and the high response rate on the sample randomly designed from the original sample. The advantage of this research is that it embedded the vignette study in a survey of a representative sample of all German companies, which is also completed by business owners or hiring managers, very specific and ideal groups for the purposes of the study.

Other more classic remedies for the problems of external validity affecting vignette studies can be cited, solutions that are not exclusive to this technique. One of them consists of replicating the studies with other samples in the same or other countries and, if we take recruitment as an example, choosing other sectors of activity. If there is a body of research that obtains similar results, the more generalizable the conclusions.

Steiner et al. (2016: 76) went one step further and pointed out that those who do not trust the external validity of any research, not only of vignette studies, but also of non-probability samples, may disdain statistical inferences and read the results simply as the estimates of the descriptive parameters of the sample. Taken to extreme skepticism, it can be concluded that each finding

is externally valid, but only under a given set of conditions (Keuschigg and Wolbring, 2016), in other words, the results are valid for the subjects analyzed in the conditions under which they have been analyzed.

Despite the doubts surrounding the extent to which the results of vignette studies or of any experiment not applied to a representative sample can be generalized, it can finally be argued that this type of design contributes to increasing sociological knowledge, since, from the perspective of some authors such as Falk and Heckman (2009), the fundamental task of the social sciences is to know the causal effect between social phenomena. That is, what is discovered has strong internal validity or what is said is true. In this sense, if we speak of certainty about causal relationships, we must agree that experimental or quasi-experimental designs win the day.

5. Conclusions

Factorial surveys, often referred to as vignette studies, are a type of quasi-experimental, survey-based study that have two main advantages. The first is that, by means of an indirect approach to the question (through the presentation of a fictitious situation), they are more suitable than traditional surveys for assessing prejudices, stereotypes or other attitudes subject to problems of social desirability. The second refers to the fact that by manipulating the fictitious scenarios presented, respondents evaluate combinations of variables that rarely occur in real-life situations. Therefore, this technique has frequently been used to evaluate labor market preferences or discrimination. Although this technique is usually applied to general populations, very specific groups such as personnel recruiters are also often analyzed, which conditions the external validity of the findings, since their participation in vignette studies is not usually based on representative sampling. It should be stressed that this limitation applies to any survey aimed at a very specific population with sampling problems. Nonetheless, the use of vignette studies is a creative and inexpensive technique that, *a priori*, presents an enormous versatility for identifying opinions, attitudes and even potential discriminatory behaviors.

Acknowledgments

We thank the two anonymous reviewers for their valuable comments, which have greatly improved the final version of the article.

Funding

This article forms part of the research projects of the National I+D+I Plan funded by the Spanish Ministry of Economy and Competitiveness with references PID2020-114702RA-I00 and PDI2020-119339GB-C21; and of the project “Men in Care: Workplace Support for Caring Masculinities” (VS/2018/0417) funded by the European Commission.

Bibliographic references

- AUSPURG, Katrin and HINZ, Thomas (2015). *Factorial Survey Experiments*. Los Angeles: Sage Series Quantitative Applications in the Social Sciences, n°. 175.
<<https://doi.org/10.4135/9781483398075>>
- AUSPURG, Katrin and JÄCKLE, Annette (2015). “First Equals Most Important? Order Effects in Vignette-Based Measurement”. *Sociological Methods & Research*.
<<https://doi.org/10.1177/0049124115591016>>
- BEEK, Krijn van; KOOPMANS, Carl and PRAAG, Bernard van (1997). “Shopping at the labour market: A real tale of fiction”. *European Economic Review*, 41, 295-317.
<[https://doi.org/10.1016/S0014-2921\(96\)00037-2](https://doi.org/10.1016/S0014-2921(96)00037-2)>
- BENARD, Stephen and CORRELL, Shelley J. (2010). “Normative Discrimination and the Motherhood Penalty”. *Gender & Society*, 24, 616-46.
<<https://doi.org/10.1177/0891243210383142>>
- BIESMA, Regien; PAVLOVA, Milena; MERODE, Frits van and GROOT, Wim (2007). “Using Conjoint Analysis to Estimate Employers Preferences for Key Competencies of Master Level Dutch Graduates Entering the Public Health Field”. *Economics of Education Review*, 26 (3), 375-386.
<<https://doi.org/10.1016/j.econedurev.2006.01.004>>
- BOOTH, Alison and LEIGH, Andrew (2010). “Do employers discriminate by gender? A field experiment in female-dominated occupations”. *Economics Letters*, 107 (2), 236-238.
<<https://doi.org/10.1016/j.econlet.2010.01.034>>
- CASTILLA, Emilio J. and BERNARD, Stephen (2010). “The Paradox of Meritocracy in Organizations”. *Administrative Science Quarterly*, 55 (6), 543-576.
<<https://doi.org/10.3905/JOI.2010.19.1.032>>
- CIS (2019). *La encuesta social europea 9, sesión de formación*. Centro de Investigaciones Sociológicas, octubre de 2019, presentación disponible en <https://www.cis.es/cis/export/sites/default/-Archivos/Formacion/BRIEFING_ESE9.pdf>.
- CORRELL, Shelley J.; BENARD, Stephen and PAIK, In (2007). “Getting a Job: Is There a Motherhood Penalty?”. *American Journal of Sociology*, 112 (5), 1.297-1.339.
<<https://doi.org/10.1086/511799>>
- CORTINA, Clara; RODRÍGUEZ, Jorge and GONZÁLEZ, M. J. (2021). “Mind the Job: The Role of Occupational Characteristics in Explaining Gender Discrimination”. *Social Indicators Research*, 156, 91-110.
<<https://doi.org/10.1007/s11205-021-02646-2>>
- CUDDY, Amy J. C.; FISKE, Susan T. and GLICK, Peter (2004). “When Professionals Become Mothers, Warmth Doesn’t Cut the Ice”. *Journal of Social Issues*, 60 (4), 701-718.
<<https://doi.org/10.1111/j.0022-4537.2004.00381.x>>
- DAMELANG, Andreas and ABRAHAM, Martin (2016). “You Can Take Some of It with You! A Vignette Study on the Acceptance of Foreign Vocational Certificates and Ethnic Inequality in the German Labour Market”. *Zeitschrift Für Soziologie*, 45 (2), 91-106.
<<https://doi.org/10.1515/zfsoz-2015-1005>>
- DI STASIO, Valentina (2014a). “Education as a Signal of Trainability: Results from a Vignette Study with Italian Employers”. *European Sociological Review*, 30 (4), 796-809.
<<https://doi.org/10.1093/esr/jcu074>>

- (2014b). *Why education matters to employers: a vignette study in Italy, England and the Netherlands*. Amsterdam: University of Amsterdam.
- DI STASIO, Valentina and GËRKHANI, Klarita (2015). "Employers' Social Contacts and Their Hiring Behaviour in a Factorial Survey". *Social Science Research*, 51, 93-107. <<https://doi.org/10.1016/j.ssresearch.2014.12.015>>
- DI STASIO, Valentina and WERFHORST, Herman G. van de (2016). "Why Does Education Matter to Employers in Different Institutional Contexts? A Vignette Study in England and the Netherlands". *Social Forces*, 95 (1), 77-106. <<https://doi.org/10.1093/sf/sow027>>
- DÜLMER, Hermann (2007). "Experimental Plans in Factorial Surveys. Random or Quota Design?". *Sociological Methods & Research*, 35 (3), 1-28. <<https://doi.org/10.1177/0049124106292367>>
- EIFLER, Stefanie and PETZOLD, Knut (2019). "Validity Aspects of Vignette Experiments: Expected 'What-If Differences Between Reports of Behavioral Intentions and Actual Behavior". In: LAVRAKAS, Paul J.; TRAUOGOTT, Michael W.; KENNEDY, Courtney; HOLBROOK, Allyson L.; LEEUW, Edith D. de and WEST, Brady T. (eds.). *Experimental Methods in Survey Research: Techniques that Combine Random Sampling with Random Assignment*. John Wiley & Sons, Inc. <<https://doi.org/10.1002/9781119083771.ch20>>
- ESCOT, Lorenzo; BELOPE-NGUEMA, Sabina; FERNÁNDEZ-CORNEJO, José Andrés; POZO-GARCÍA, Eva del; CASTELLANOS-SERRANO, Cristina and CRUZ-CALDERÓN, Selene Fabiola (2021). "Can the legal framework for prostitution influence the acceptability of Buying sex?". *Journal of Experimental Criminology*. <<https://doi.org/10.1007/s11292-021-09465-y>>
- FALK, Armin and HECKMAN, James J. (2009). "Lab Experiments are a Major Source of Knowledge in the Social Sciences". *Science*, 326, 535-538. <<https://doi.org/10.1126/science.1168244>>
- FERNÁNDEZ-LOZANO, Irina; GONZÁLEZ, M. José; JURADO-GUERRERO, Teresa and MARTÍNEZ-PASTOR, Juan-Ignacio (2020). "The Hidden Cost of Flexibility: A Factorial Survey Experiment on Job Promotion". *European Sociological Review*, 36 (2), 265-283. <<https://doi.org/10.1093/esr/jcz059>>
- GADDIS, Michael S. (2015). "Discrimination in the credential society: An audit study of race and college selectivity in the labor market". *Social Forces*, 93 (4), 1.451-1.459. <<https://doi.org/10.1093/sf/sou111>>
- GERRING, John (2014). *Metodología de las ciencias sociales*. Madrid: Alianza.
- GNIZA, Jan; MARTIN, Abraham; LEHMAN, Isabella and LOREK, Kerstin (2022). "Expecting equality or efficiency? A factorial survey on the justive perception of regional redistributive policies". *Regional Studies*. <<https://doi.org/10.1080/00343404.2021.2018129>>
- GUTFLEISCH, Tamara; ROBIN, Samuel and SACCHI, Stefan (2021). "The application of factorial surveys to study recruiters' hiring intentions: comparing designs based on hypothetical and real vacancies". *Quality & Quantity*, 55, 775-804. <<https://doi.org/10.1007/s11135-020-01012-7>>
- HAINMUELLER, Jens; HANGARTNER, Dominik and YAMAMOTO, Teppei (2015). "Validating vignette and conjoint survey experiments against real-world behavior". *PNAS*, 112 (8), 2.395-2.400. <<https://doi.org/10.1073/pnas.1416587112>>

- HENKENS, Kène; SOLINGE, Hanna van and COZIJNSEN, Rabina (2009). "Let Go or Retain? A Comparative Study of the Attitudes of Business Students and Managers about the Retirement of Older Workers". *Journal of Applied Social Psychology*, 39 (7), 1.562-1.588.
<<https://doi.org/10.1111/j.1559-1816.2009.00495.x>>
- HUMBURG, Martin and VELDEN, Rolf van der (2014). *Skills and the Graduate Recruitment Process: Evidence From Two Discrete Choice Experiments*. Research Centre for Education and the Labour Market.
<<https://doi.org/10.1016/j.econedurev.2015.07.001>>
- JACOBS, Jerry A. and GERSON, Kathleen (2016). "Unpacking americans' views of the employment of mothers and fathers using national vignette survey data: SWS Presidential Address". *Gender & Society*, 30 (3), 413-441.
<<https://doi.org/10.1177/0891243215597445>>
- KARPINSKA, Kasia; HENKENS, Kène and SCHIPPERS, Joop (2011). "The Recruitment of Early Retirees: A Vignette Study of the Factors That Affect Managers' Decisions". *Ageing and Society*, 31 (4), 570-589.
<<https://doi.org/10.1017/S0144686X10001078>>
- (2013). "Retention of Older Workers: Impact of Managers' Age Norms and Stereotypes". *European Sociological Review*, 29 (6), 1.323-1.335.
<<https://doi.org/10.1093/esr/jct017>>
- KARPINSKA, Kasia; HENKENS, Kène; SCHIPPERS, Joop and WANG, Mo (2015). "Training Opportunities for Older Workers in the Netherlands: A Vignette Study". *Research in Social Stratification and Mobility*, 41, 103-112.
<<https://doi.org/10.1016/j.rssm.2015.03.002>>
- KARREN, Ronald J. and BARRINGER, Melissa Woodbard (2022). "A Review and Analysis of the Policy-Capturing Methodology in Organizational Research: Guidelines for Research and Practice". *Organizational Research Methods*, 5 (4), 337-361.
<<https://doi.org/10.1177/109442802237115>>
- KENNEDY, Courtney and HARTIG, Hannah (2019). "Response rates in telephone surveys have resumed their decline". *Pew Research Center*.
<<https://pewrsr.ch/2XqsgTT>>
- KEUSCHNIGG, Marc and WOLBRING, Tobias (2016). "The Use of Field Experiments to Study Mechanisms of Discrimination". *Analyse & Kritik. Zeitung für linke Debatte und Praxis*, 38 (1), 179-202.
<<https://doi.org/10.1515/auk-2016-0109>>
- KING, Gary; KEOHANE, Robert O. and VERBA, Sidney (2005). *El diseño de la investigación social. La inferencia científica en los estudios cualitativos*. Madrid: Alianza.
- KOOTSTRA, Anouk (2016). "Deserving and Undeserving Welfare Claimants in Britain and the Netherlands: Examining the Role of Ethnicity and Migration Status Using a Vignette Experiment". *European Sociological Review*, 32 (3), 325-338.
<<https://doi.org/10.1093/esr/jcw010>>
- KÜBLER, Dorothea; SCHMID, Julia and STÜBER, Robert (2018). "Gender discrimination in hiring across occupations: a nationally-representative vignette study". *Labour Economics*, 55, 215-229.
<<https://doi.org/10.1016/j.labeco.2018.10.002>>
- LAENEN, Tigs; HOOTEGEM, Arno van and ROSSETTI, Federica (2022). "The multidimensionality of public support for basic income: a vignette experiment in Belgium". *Journal of European Public Policy, online*.
<<https://doi.org/10.1080/13501763.2022.2055112>>

- LIECHTI, Fabienne; FOSSATI, Flavia; BONOLI, Giuliano and AUER, Daniel (2017). "The Signalling Value of Labour Market Programmes". *European Sociological Review*, 33 (2), 257-274.
<<https://doi.org/10.1093/esr/jcw061>>
- MARTÍNEZ PASTOR, Juan Ignacio (2019). *Los datos sin tapujos. Cómo interpretar y difundir las estadísticas sociales*. Madrid: Los Libros de la Catarata.
- MCDONALD, Patrick (2019). "How Factorial Survey Analysis Improves Our Understanding of Employer Preferences". *Swiss Journal of Sociology*, 45 (2), 237-260.
<<https://doi.org/10.2478/sjs-2019-0011>>
- MULDERS, Jaap; DALEN, Hendrik van; HENKENS, Kène and SCHIPPERS, Joop (2014). "How Likely Are Employers to Rehire Older Workers after Mandatory Retirement? A Vignette Study Among Managers". *De Economist*, 162, 415-431.
<<https://doi.org/10.1007/s10645-014-9234-8>>
- OESCH, Daniel; LIPPS, Oliver and MCDONALD, Patrick (2017). "The Wage Penalty for Motherhood: Evidence on Discrimination From Panel Data and a Survey Experiment for Switzerland". *Demographic Research*, 37, 1.793-1.824.
<<https://doi.org/10.4054/DemRes.2017.37.56>>
- RUDMAN, Laurie A. and MESCHER, Kris (2013). "Penalizing men who request a family leave: Is flexibility stigma a femininity stigma?". *Journal of Social Issues*, 69 (2), 322-340.
<<https://doi.org/10.1111/josi.12017>>
- SHI, Lulu P.; IMDORF, Christian; SAMUEL, Robin and SACCHI, Stefan (2018). "How Unemployment Scarring Affects Skilled Young Workers: Evidence From a Factorial Survey of Swiss Recruiters". *Journal for Labour Market Research*, 52 (7), 1-15.
<<https://doi.org/10.1186/s12651-018-0239-7>>
- STEINER, Peter, M.; ATZMÜLLER, Christiane and SU, Dan (2016). "Designing Valid and Reliable Vignette Experiments for Survey Research: A Case Study on the Fair Gender Income Gap". *Journal of Methods and Measurement in the Social Sciences*, 7 (2), 52-94.
<<https://doi.org/10.2458/jmm.v7i2.20321>>
- STERKENS, Philippe; BAERT, Stijn; ROOMAN, Claudia and DEROUS, Eva (2020). "As If it Weren't Hard Enough Already: Breaking Down Hiring Discrimination Following Burnout". *IZA Institute of Labor Economics, Discussion Paper Series No. 13514*.
<<https://doi.org/10.2139/ssrn.3660260>>
- WALLANDER, Lisa (2009). "25 years of factorial surveys in sociology: A review". *Social Science Research*, 38 (3), 505-520.
<<https://doi.org/10.1016/j.ssresearch.2009.03.004>>
- WOLF, Inge de and VELDEN, Rolf van der (2001). "Selection Processes for Three Types of Academic Jobs: An Experiment Among Dutch Employers of Social Sciences Graduates". *European Sociological Review*, 17 (3), 317-330.
<<https://doi.org/10.1093/esr/17.3.317>>

