

*The application of factorial surveys to study  
recruiters' hiring intentions: comparing  
designs based on hypothetical and real  
vacancies*

**Tamara Gutfleisch, Robin Samuel &  
Stefan Sacchi**

**Quality & Quantity**  
International Journal of Methodology

ISSN 0033-5177

Qual Quant  
DOI 10.1007/s11135-020-01012-7



**Your article is published under the Creative Commons Attribution license which allows users to read, copy, distribute and make derivative works, as long as the author of the original work is cited. You may self-archive this article on your own website, an institutional repository or funder's repository and make it publicly available immediately.**



# The application of factorial surveys to study recruiters' hiring intentions: comparing designs based on hypothetical and real vacancies

Tamara Gutfleisch<sup>1</sup> · Robin Samuel<sup>1</sup> · Stefan Sacchi<sup>2</sup>

© The Author(s) 2020

## Abstract

Factorial survey experiments have been widely used to study recruiters' hiring intentions. Respondents are asked to evaluate hypothetical applicant descriptions, which are experimentally manipulated, for hypothetical job descriptions. However, this methodology has been criticized for putting respondents in hypothetical situations that often only partially correspond to real-life hiring situations. It has been proposed that this criticism can be overcome by sampling real-world vacancies and the recruiters responsible for filling them. In such an approach, only the applicants' descriptions are hypothetical; respondents are asked about a real hiring problem, which might increase internal and external validity. In this study, we test whether using real vacancies triggers more valid judgments compared to designs based on hypothetical vacancies. The growing number of factorial survey experiments conducted in employer studies makes addressing this question relevant, both for methodological and practical reasons. However, despite the potential implications for the validity of data, it has been neglected so far. We conducted a factorial survey experiment in Luxembourg, in which respondents evaluated hypothetical applicants referring either to a currently vacant position in their company or to a hypothetical job. Overall, we found little evidence for differences in responses by the design of the survey experiment. However, the use of real vacancies might prove beneficial depending on the research interest. We hope that our comparison of designs using real and hypothetical vacancies contributes to the emerging methodological inquiry on the possibilities and limits of using factorial survey experiments in research on hiring.

**Keywords** Factorial survey experiments · Hiring intentions · Real vacancies · Hypothetical vacancies

---

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1007/s11135-020-01012-7>) contains supplementary material, which is available to authorized users.

---

✉ Tamara Gutfleisch  
tamara.gutfleisch@uni.lu

<sup>1</sup> Department of Social Sciences, University of Luxembourg, 2 avenue de l'Université, 4365 Esch-sur-Alzette, Luxembourg

<sup>2</sup> Institute of Sociology, University of Bern, Fabrikstrasse 8, 3012 Bern, Switzerland

# 1 Introduction

Factorial survey experiments (FSEs) have been used in the social sciences to study human judgments since the 1970s (Rossi 1979; Wallander 2009) and, in recent years, have become an established method to study how employers select job applicants (McDonald 2019). In a typical set-up, respondents are asked to evaluate fictitious descriptions of applicants, called “vignettes.” Researchers can study how employers interpret and value specific applicant characteristics during the hiring process through the experimental manipulation of information on applicants. Analyzing employers’ hiring preferences is crucial to our understanding of how inequalities arise at the individual and job level (Bills et al. 2017). Previous studies have employed FSEs to analyze, for example, gender discrimination (Kübler et al. 2018; Oesch et al. 2017), ethnic discrimination (Auer et al. 2019; Baert and De Pauw 2014), the value of foreign qualifications (Damelang et al. 2019; Protsch and Solga 2017), and the signaling value of educational credentials (Di Stasio 2014; Di Stasio and Van De Werfhorst 2016) and unemployment experience (Van Belle et al. 2018; Shi et al. 2018).

However, FSEs have been criticized for measuring the hiring intentions of employers, rather than their real behavior (e.g., Pager and Quillian 2005). Consequently, field experiments, such as correspondence or audit studies in which fictitious applications are sent to real jobs (e.g., Pedulla 2018; Protsch and Solga 2014), are often seen as better suited to detecting inequalities in the hiring process.<sup>1</sup>

This criticism applies particularly to FSEs, in which respondents are put in a situation that may be completely unknown to them in real life. This would be the case, for example, if respondents are students (e.g., Baert and De Pauw 2014) or only partly familiar with the recruitment process for a specific job type (e.g., Van Belle et al. 2018; Damelang and Abraham 2016). However, FSEs can also be based on a careful selection of study participants, so that the pool of respondents closely matches the target population (i.e., individuals with recruiting experience for the occupation of interest and familiar with the situations described in the FSE). Adopting this strategy may increase external validity (Hainmueller et al. 2015).

Furthermore, FSEs have several advantages over field experiments, which make them an attractive option for studying hiring processes (see also, McDonald 2019). First, unlike field experiments, FSEs may capture the multidimensionality of hiring decisions by considering several applicant characteristics simultaneously instead of focusing on only one or two features. Moreover, the survey part of FSEs allows the researcher to gather rich information about recruiters, companies, and jobs; this enables an in-depth analysis of the mechanisms underlying recruiters’ decision-making processes, an analytical feature typically unavailable in field experiments. Finally, FSEs are easily compatible with ethical research practices, as the respondents (i.e., recruiters) give informed consent to participate in the survey, whereas in field experiments, real employers are led to believe that they are dealing with real applications.<sup>2</sup>

Resolving the FSE versus field experiment debate is beyond the scope of this article. Given that prior research suggests that FSEs are useful, some of the justified criticisms about their use should be addressed by improving the design of the FSEs applied to study hiring intentions. A research group from the NEGOTIATE project proposed basing FSEs on samples of current real-world vacancies and the recruiters responsible for filling them (NEGOTIATE

<sup>1</sup> See Baert (2018) for an overview of field experiments to study hiring decisions.

<sup>2</sup> See Keuschnigg and Wolbring (2016) and Riach and Rich (2004) for a detailed discussion of the possible real-life impact of this deception on employers and applicants.

2020).<sup>3</sup> This is in contrast to most previous FSEs on recruiting, which relied on rather abstract descriptions of hypothetical vacancies that may have lacked psychological realism (e.g., Van Belle et al. 2018; Damelang et al. 2019; Di Stasio 2014). “*Psychological realism* [emphasis in original] is the extent to which the psychological processes that occur in an experiment are the same as the psychological processes that occur in everyday life” (Brewer and Crano 2014, p. 21). This is an important factor that should be considered when designing FSEs, to enhance the external validity (Auspurg and Hinz 2015). Conducting FSEs with recruiters who are currently in charge of filling a real-world vacancy might increase psychological realism and, hence, contribute to the internal and external validity of the results (NEGOTIATE 2020). However, collecting large sets of job advertisements and employer details is costly and time-consuming. Furthermore, given that respondents are still presented with a fictitious situation that entails no real-life consequences, the question arises whether it really makes a difference for the measured hiring intentions of recruiters whether vignette ratings refer to a real hiring problem or a hypothetical one. If the FSE is based on hypothetical vacancies, the employers might have trouble putting themselves in the actual hiring situation, particularly when the respondents are not familiar with recruiting for the specific job type studied. Hence, the internal and external validity might be reduced. This is presumably different if the FSE is conducted with recruiters who are currently responsible for filling a real-world vacancy on which the experiment is based. Only the applicant profiles are fictitious in this case, which might increase internal and external validity. However, the question remains whether this is enough to trigger more valid judgments compared to a completely hypothetical situation.

The main objective of this study is to help answer this question. While the advantages and disadvantages of survey experiments in general have been discussed (e.g., Sniderman 2018), more research is needed on the designs of FSEs to study hiring intentions. We conducted an FSE in Luxembourg, where one group of recruiters rated hypothetical applicants for a real vacancy for which they were personally responsible in real life (i.e., for which they know all the requirements and characteristics) and another group of recruiters rated hypothetical applicants for a hypothetical vacancy (i.e., without information on job characteristics). Since random assignment of recruiters to a condition with real or hypothetical vacancies (i.e., a split ballot experiment) was impossible due to data limitations, we compared two different samples of recruiters. We manipulated three applicant characteristics: migrant background, gender, and unemployment. We asked whether the effects of these characteristics on observed hiring intentions differed if the rating task corresponded to a real instead of a hypothetical hiring problem and whether these effects were closer to certain benchmarks when real vacancies were used. “Behavioral benchmarks” (see Hainmueller et al. 2015) corresponding to real-life hiring behavior were not available for our study context, which is why we derived “theoretical” benchmarks from previous research. Given the growing number of FSEs in employer studies and the potential implications for the validity of those studies’ results, answering this question is relevant for both methodological and practical reasons. However, to the best of our knowledge, this question has not been addressed empirically.

After a literature review in Sect. 2, we discuss our expectations about the impact of the FSE design on recruiters’ stated hiring intentions in Sect. 3. In Sect. 4, we describe our research design. Our results are presented in Sect. 5 Section 6 contains our conclusions.

---

<sup>3</sup> Within the context of the Horizon 2020 project NEGOTIATE, a multinational FSE was conducted in four European countries, in which the rating task of hypothetical applicants was based on current real-world vacancies the respondents were responsible for filling. The data was used in several scientific publications (Bertogg et al. 2020; Imdorf et al. 2019; Shi et al. 2018).

## 2 Literature review

Table 1 provides an overview of previous FSEs used to study recruiters' hiring intentions.<sup>4</sup> All listed studies ask respondents to rate either the hiring chances of applicants or the likelihood of inviting applicants to the next selection stage (e.g., interview). Moreover, in all cases, the presented applicant profiles (i.e., vignettes) are fictitious. These studies concern how recruiters interpret and value certain applicant characteristics, such as educational background, age, or nationality.

The last column of Table 1 shows the type of jobs presented to respondents. To the best of our knowledge, so far only in NEGOTIATE (2020) have respondents been asked about a real vacancy in their company that they are responsible for filling.<sup>5</sup> Scholars interested in measuring hiring intentions by means of FSEs clearly prefer using hypothetical vacancies. However, the presentation of hypothetical jobs to respondents varies in certain ways. For example, some studies give specific information about the job, such as characteristics required for the position (e.g., Baert and De Pauw 2014; De Wolf and Van Der Velden 2001); other studies are more general and provide no details about the job (e.g., Oesch et al. 2017). Furthermore, some rating tasks do not refer to a job *per se* but to the organization as a whole (e.g., Karpinska et al. 2011).

The pool of respondents used in the reviewed studies also differs noticeably. The respondents in most studies are managers or human resource professionals, who have at least some recruitment experience (see the 5th column in Table 1). Yet, they are not always experts in the specific occupations under investigation or in the relevant recruitment procedures (e.g., Damelang and Abraham 2016). In a few cases, the sample consists fully or partly of students (Baert and De Pauw 2014; Karpinska et al. 2011). Such a mismatch between the target population (i.e., real recruiters with job-specific knowledge) and the analytic sample could affect the internal and external validity of the results, as discrepancies between measured intentions and real behavior are more likely if respondents are not familiar with the rating task in real life (see Todesco 2017, p. 17). Hence, FSEs based on hypothetical vacancies can enhance the validity of the results by carefully selecting respondents with relevant recruiting experience (ideally for the types of jobs under investigation) and designing hypothetical vacancies that are familiar to them. Nevertheless, this issue can be easily addressed if the sample consists of real-world vacancies, given that, in this case, the job described in the hiring situation is concrete and real for the respondents. The respondents (i.e., real recruiters) are asked about a currently vacant position in their company, for which they are *personally* responsible. This means that they know all the relevant characteristics of this vacancy and do not have to make any assumptions when rating vignettes (see Sect. 3). This is different for respondents who

<sup>4</sup> McDonald (2019) provides a more general systematic review of previous FSEs used in the study of employers' preferences. In our literature review, we compare previous FSEs based on their use of hypothetical and real-world vacancies. We focus on FSEs published in peer-reviewed journals written in English and exclude behavioral validation studies (i.e., studies that combine field and survey experiments to compare the two methods). Conjoint analyses and other types of vignette studies, which are similar to FSEs but typically ask respondents to rate applicant profiles in pairs rather than individually (e.g., Auer et al. 2019; Van Beek et al. 1997; Biesma et al. 2007; Humburg and Van Der Velden 2015), are also excluded as our study design diverges from this approach.

<sup>5</sup> Van Beek et al. (1997) also asked employers about a real vacancy in their company in their FSE, but they adopted a different approach. They asked a sample of employers to *describe* a real (current or past) vacancy in their firm, and the vignette ratings were based on this description. In contrast, NEGOTIATE (2020) sampled published real vacancies, and the study participants were personally responsible for filling them. Moreover, Van Beek et al. (1997) asked respondents to rate pairs of applicants rather than rate the applicants in consecutive order.

**Table 1** Overview of FSE about hiring intentions

Study	Subject	Country	Data/sample	Respondents	Familiar with occupation	Job type
Oesch et al. (2017)	Discrimination (Gender/motherhood)	CH	Members from human resources management association	Human resource managers	No/not necessarily	Hypothetical vacancy
Bertogg et al. (2020)	Discrimination (Gender)	Four countries	NEGOTIATE, sample of real vacancies	Persons responsible for filling a particular vacancy	Yes	Real vacancy the respondent is personally responsible for filling
Kübler et al. (2018)	Discrimination (Gender)	DE	BIBB, Company Panel on Qualification and Competence Development 2014	Firm owners and human resource managers	No/not necessarily	Hypothetical apprenticeship position
Pedulla (2016)	Discrimination (Gender)	USA	Online Opt-In Panel	Individuals who make hiring decisions	No/not necessarily	Hypothetical vacancy
Baert and De Pauw (2014)	Discrimination (nationality, ethnicity)	BE	Undergraduates at Ghent University	Students in economics	No/not necessarily	Hypothetical vacancy; some details about job requirements
Di Stasio (2014)	Education/trainability	IT	Random sample of organizations	Recruiters and human resource professionals	Yes	Hypothetical vacancies; detailed job description
Di Stasio (2017)	Education/trainability	UK & NL	Random sample of organizations	Recruiters and human resource professionals	Yes	Hypothetical job description; detailed job description



Table 1 continued

Study	Subject	Country	Data/sample	Respondents	Familiar with occupation	Job type
Di Stasio and Gërxhani (2015)	Education/trainability	UK	Respondents randomly sampled from association records	Human resource professionals	Yes	Hypothetical vacancy; detailed job description
Di Stasio and Van De Werfhorst (2016)	Education/trainability	UK & NL	Random sample of organizations based on public records	Recruiters and human resource professionals	Yes	Hypothetical vacancy; detailed job description
Wilson (2019)	Academic achievement	CH	Registry of in-firm vocational instructors	Vocational trainers active in commercial training	Yes	Hypothetical apprenticeship position
De Wolf and Van Der Velden (2001)	Hiring for high-skilled jobs	NL	Alumni from Faculty of Social Sciences at Utrecht University	Recruiters	Yes	Hypothetical vacancy; basic description of job
Damelang and Abraham (2016)	Immigration/foreign qualification	DE	Cluster-Oriented Regional Information System (CORIS) data base	90% of sample responsible for personnel decisions	No/not necessarily	Hypothetical vacancy
Damelang et al. (2019)	Immigration/foreign qualification	DE	Organizations offering apprenticeships	93% of respondents responsible for personnel decisions	Yes	Hypothetical vacancy
Damelang et al. (2020)	Immigration/foreign qualification	DE	Employment history data from social security notifications	Managers who are responsible for hiring decisions	Yes	Hypothetical vacancy
Mergener and Maier (2018)	Immigration/foreign qualification	DE	BIBB Recognition Monitoring, German employer survey	Human resource managers	No/not necessarily	Hypothetical vacancy
Protsch and Solga (2017)	Immigration/foreign qualification	DE	BIBB Training Panel 2014 (representative sample of all firms in Germany)	Individuals involved in human resource activities	Yes	Hypothetical apprenticeship position



Table 1 continued

Study	Subject	Country	Data/sample	Respondents	Familiar with occupation	Job type
Karpinska et al. (2011)	Hiring older workers	NL	The Netherlands Interdisciplinary Demographic Institute/Utrecht University	Managers and business students	-	No specific position; referring to overall organization
Liechti et al. (2017)	Participation in labor market programs	CH	Sample drawn from Swiss hotel employer association	Hotel managers	Yes	Hypothetical vacancy
Van Belle et al. (2019)	Participation in labor market programs	BE	Database of individuals living in Flanders	Professionals familiar with real-life hiring processes	No/not necessarily	Hypothetical vacancy at fictitious company
Van Belle et al. (2018)	Unemployment	BE	Database of individuals living in Flanders	Professionals familiar with real-life hiring processes	No/not necessarily	Hypothetical vacancy at fictitious company
Shi et al. (2018)	Unemployment	CH	NEGOTIATE, sample of real vacancies	Persons responsible for filling a particular vacancy	Yes	Real vacancy the respondent is personally responsible for filling
Petzold (2017)	Student mobility	DE	Sample of real vacancies	Individuals involved in recruitment for vacancies	No/not necessarily	Hypothetical vacancy within organization

may have relevant recruiting experience but are presented with a rather vague description of a hypothetical vacancy.

Further differences among FSEs used to measure hiring intentions involve the type of samples used. Respondents were either drawn from a representative population or employer surveys or were sampled specifically for the respective study context (see column 4 in Table 1). For example, Liechti et al. (2017) sampled respondents based on the membership register of a Swiss hotel employer association, whereas other studies used a (random) sample of firms to find suitable study participants (e.g., Van Beek et al. 1997; Di Stasio 2014). Similar to the NEGOTIATE research group, Petzold (2017) sampled real job advertisements through which recruiters were contacted. As far as we can judge, however, Petzold (2017), in contrast to NEGOTIATE (2020), presented respondents with a hypothetical job description.

Some of these differences may be due to the different study contexts and goals, but they also indicate an evolving field that has not yet converged upon shared standards. Whereas the vast majority of the reviewed studies arrive at conclusions in line with the respective theory, common guidelines will increase their comparability.

### 3 Theory and hypotheses

Our contribution focuses on the applicant's migrant background, gender, and unemployment experience to examine whether the FSE design affects recruiters' hiring intentions. Hiring-related inequalities based on these characteristics have been of particular interest to scholars in labor market research. In the following, we develop the argument that the effects of these characteristics on recruiters' hiring intentions are closer to what we would observe in reality when real vacancies are used. To test this hypothesis, behavioral benchmarks corresponding to the true effects of these characteristics on hiring decisions would be ideal (see Hainmueller et al. 2015), but these were not available for our study context. Therefore, we had to rely on the extensive empirical evidence on the effects of our three applicant characteristics on recruiters' hiring decisions to find plausible "theoretical" benchmarks. These will be briefly discussed below, before we turn to our expectations about the impact of using real vacancies in FSEs on recruiters' hiring intentions.

In line with well-established discrimination theories, hiring discrimination toward foreigners and ethnic minorities has been documented through numerous field (e.g., Zschirnt and Ruedin 2016; Quillian et al. 2019) and survey experiments (e.g., Auer et al. 2019). Therefore, we assume a negative effect of a foreign background on recruiters' hiring intentions as the benchmark for the effect of migrant background for our study context. Moreover, experimental studies suggest an advantage for women in female-dominated occupations and a disadvantage for women in male-dominated occupations (e.g., Birkelund et al. 2019; Koch et al. 2015; Fernandez and Mors 2008). Based on this evidence, we assume two benchmarks for the effect of gender for our study context: a negative (positive) effect of being female in male-dominated (female-dominated) occupations. However, the empirical evidence on the effect of gender on recruitment decisions is somewhat mixed, and these benchmarks might therefore be less reliable. For example, some studies contest the notion of hiring discrimination towards women in male-dominated occupations (e.g., Petersen and Togstad 2006; Carlsson 2011). Finally, with only a few exceptions (e.g., Nunley et al. 2017), several experimental studies have shown that employers are more reluctant to hire applicants who have been unemployed (e.g., Van Belle et al. 2018; Eriksson and Rooth 2014; Kroft et al. 2013;

Oberholzer-Gee 2008). Following this literature, we assume a negative effect of unemployment on recruiters' hiring intentions as our benchmark.

We explore why and how these effects might differ when the rating task refers to a real hiring problem as opposed to a hypothetical one. First, the hypothetical job descriptions and vignettes presented in prior FSEs likely differ from real-life hiring situations, which might lead to a "hypothetical bias" (Ajzen et al. 2004), meaning that respondents answer differently than they would in reality. However, the psychological realism of the rating task increases decisively when the FSE is based on real vacancies for which the respondents are *personally* responsible in real life. In this case, the vacancies are familiar to the respondents, and they can draw on the same information (e.g., regarding requirements) they would normally do to form hiring decisions for this particular job type. Although hypothetical vacancies might allow researchers to control for possible differences in job requirements, it is difficult to determine whether the information provided in fictitious job descriptions will sufficiently chime with the psychological realism held by respondents. A careful design, such as asking recruiters about typical vacancies in their company (e.g., Van Beek et al. 1997; Di Stasio 2014), could alleviate the potential problem of low psychological realism. Nevertheless, respondents must rely on assumptions about requirements and characteristics of the respective hypothetical vacancy on which the experiment is based, which may differ from real vacancies. The latter is all the more likely the less familiar the respondents are with typical vacancies in the respective occupation studied and the less specific the descriptions of these vacancies are. In contrast, while vignettes only represent simplified applicant profiles (i.e., they contain less information than real application documents), they usually provide realistic information about applicants according to the researchers' interest (if carefully designed).<sup>6</sup> This holds regardless of whether real or hypothetical vacancies are used. Yet, respondents who are presented with a real vacancy will have less difficulty deciding whether the provided information corresponds to reality and matches the respective job characteristics. As a corollary, the information respondents consider when rating the vignettes might be closer to real-life decisions made throughout the hiring process. Therefore, the main effects of gender, migration background, and unemployment should be closer to our derived benchmarks if real vacancies are used.

Second, social desirability is a well-known problem in survey research that might bias results. Although the multidimensional design of FSEs should mitigate the risk of normative influences on vignette ratings (Auspurg et al. 2015), the possibility of social desirability bias cannot be fully excluded. Socially biased results would imply less discrimination by gender or migration background and less negative unemployment effects than theoretically expected. Recruiters might be more cautious with their answers in both FSE versions, as they cannot be sure about the real reason for the experiment; however, they might feel more confident about their vignette ratings when the FSE is based on real vacancies. This is because, as we have argued, real vacancies provide psychological realism to the rating task and recruiters may find it easier to justify their answers if needed. For similar reasons, the use of real vacancies might also reduce the risk of other response biases, such as acquiescence (Krosnick et al. 2014; Schwarz 1999), as respondents might be generally less susceptible to external influences if they are confronted with a real-life hiring problem. Consequently, in line with our considerations above, the effects of applicant characteristics should be closer to our benchmarks when real instead of hypothetical vacancies are used. However, the effects of applicant characteristics might not be affected in the same way. Reluctance to hire based on unemployment could be more easily justified with productivity-related characteristics (e.g.,

<sup>6</sup> For example, the vignettes in NEGOTIATE (2020) displayed real and common job titles and educational credentials from the respective occupation.

skill loss) than reluctance based on characteristics such as gender or migration background, where social desirability is more likely an issue.

Third, FSEs have been criticized for presenting simplified hiring situations that cannot possibly convey the urgency and pressures surrounding actual hiring (e.g., difficulty finding suitable candidates or a company's need to fill a position promptly). These aspects are more likely to be considered if respondents are confronted with a real hiring problem. Prior research suggests that gender, unemployment, or migration background might be less relevant in such a case (e.g., Baert et al. 2015). Unfortunately, we cannot analyze aspects of urgency in detail due to data limitations (see Sect. 4).

Against this background, we expect differences in the effect of applicant characteristics on recruiters' hiring intentions based on the FSE's design. More specifically, in line with our theory, we expect the effects of gender, migrant background and unemployment on vignette ratings to be closer to the defined benchmarks when real vacancies are used as compared to hypothetical vacancies (*Hypothesis 1*). Although in both cases, respondents rate hypothetical applicants without real-life consequences, we argue that the perceived realism associated with the rating task is markedly higher in the former setting. Also, the effect of unemployment might be generally less affected by social desirability than the effects of gender and migration background are. Hence, we expect the difference in the effects of gender and migrant background between the FSE designs (see *Hypothesis 1*) to be greater than the same difference regarding the effect of unemployment (*Hypothesis 2*). The hypotheses will be tested for each applicant characteristic separately.

## 4 Research design

We conducted an FSE in Luxembourg in 2018 and 2019.<sup>7</sup> To enable comparability with NEGOTIATE (2020), we built on that design using the same pictorial representation of CV and answer scale.<sup>8</sup> Respondents were asked to rate hypothetical descriptions of applicants (*vignettes*) that varied systematically on a number of different characteristics (*dimensions*). The rating task either referred to a real vacancy in their company or to a similar but hypothetical job.

The relatively small Luxembourgish labor market did not provide enough real vacancies to conduct a split ballot experiment, where we could have randomly assigned respondents to a real vacancy or a hypothetical job description in the FSE.<sup>9</sup> Therefore, we employed a two-step approach to gather our data. First, we collected real vacancies and recruiter contact information published on different online job portals and company websites in Luxembourg. Second, we sampled recruiters from publicly available lists of companies and businesses associated with the same occupational fields. Respondents sampled based on the former approach rated hypothetical applicants in the FSE based on the real vacancies they were responsible for filling. Respondents sampled based on the latter approach rated applicants for a hypothetical but similar type of job. In the present analysis, we focus on jobs in catering and

<sup>7</sup> A corresponding pilot study was conducted in spring 2018. Both the pilot and main study received approval from the Ethics Review Panel of the University of Luxembourg.

<sup>8</sup> The vignette dimensions, however, differ from the design of the NEGOTIATE survey. A pilot study showed that the NEGOTIATE design is not feasible in the context of Luxembourg, as it would require a much larger sample size to achieve enough statistical power.

<sup>9</sup> A split ballot experiment would have allowed a more accurate test of the impact of the FSE design, which we discuss in Sect. 6.

in information technology (IT).<sup>10</sup> Both groups of respondents have at least some recruitment experience in the respective occupations studied (see Sects. 4.2 and 4.3), but only respondents sampled based on real vacancies are presented with a concrete, realistic hiring problem in the FSE. This enables us to test whether the use of real vacancies triggers different responses while avoiding distortion in our results from differences in the recruitment experience between the two samples (and thus fundamental differences in perceived realism). In the following sections, we will refer to the “real vacancy” (RV) design and “hypothetical vacancy” (HV) design when discussing the two versions of the FSE. However, the reader should keep in mind that we are essentially comparing two different samples.

#### 4.1 Vignettes

The vignettes with hypothetical applicants varied in the values of three experimental variables (see Table 2)<sup>11</sup>: gender (male/female), unemployment (no unemployment/one year of unemployment after graduation/one year of current unemployment), and migrant group, which was signaled to respondents by the applicant's nationality and country of residence (Luxembourgish, foreigners: Portuguese/Luxembourgish-Portuguese/French, border workers: French/German). Luxembourgish-Portuguese nationality was used to signal dual citizenship. The migrant groups selected result from the unique demographic situation in Luxembourg, which is characterized by a multi-cultural population and workforce.<sup>12</sup> The amount of work experience was held constant in the experimental design: each vignette showed 48 months of occupation-specific work experience at a company located in Luxembourg. To make the information provided in our vignettes as realistic as possible, the professional titles all matched the most common job titles in each occupation under investigation.

The experimental design included 36 vignettes, representing all possible combinations of applicant characteristics ( $2^1 \times 3^1 \times 6^1$ ). We divided the 36 vignettes into 6 equally sized decks, each including 6 vignettes. Following Kuhfeld (2010), we used D-efficient blocking to allocate the vignettes to decks, that is, we optimized the decks for maximum variance and orthogonality between each vignette dimension (Atzmüller and Steiner 2010; Dülmer 2007). We assigned each deck randomly to respondents. As recommended in the literature (Auspurg and Hinz 2015), we randomized the order of vignettes across respondents to avoid ordering bias. For each vignette, we asked the respondents to rate the likelihood of considering the respective applicant for the advertised position in their company (RV design) or a typical position from the same field (HV design) on an 11-point scale (0 = *practically zero* to 10 =

<sup>10</sup> The FSE was actually conducted in five occupations. However, due to low sample sizes in some occupations, we are only able to consider the two occupations named in our study; we had to exclude respondents from the fields of mechanics, finance, and nursing.

<sup>11</sup> Between five and nine dimensions are usually recommended in the literature to increase the variation between vignettes and thereby avoid fatigue effects (Auspurg and Hinz 2015; Sauer et al. 2011). However, a more complex design would have required a larger sample size that would have been difficult to achieve in Luxembourg. To control for possible fatigue effects, we included indicators for the vignette order in our regression models (see Sect. 4.5).

<sup>12</sup> Almost half of the general population in 2019 were foreigners (47%), of which 33% were Portuguese, the largest foreign group, followed by the French (16%); own calculations based on data retrieved from Luxembourg's statistical office (STATEC): [https://statistiques.public.lu/stat/TableViewer/tableView.aspx?ReportId=12853&IF\\_Language=eng&MainTheme=2&FldrName=1](https://statistiques.public.lu/stat/TableViewer/tableView.aspx?ReportId=12853&IF_Language=eng&MainTheme=2&FldrName=1), January 4, 2020. Moreover, a large proportion of workers (46%) are border workers living in one of the neighboring countries but working in Luxembourg, of which French border workers make up over half (53%), followed by Germans (24%); own calculations based on data retrieved from STATEC: [https://statistiques.public.lu/stat/TableViewer/tableView.aspx?ReportId=12919&IF\\_Language=eng&MainTheme=2&FldrName=3&RFPPath=92](https://statistiques.public.lu/stat/TableViewer/tableView.aspx?ReportId=12919&IF_Language=eng&MainTheme=2&FldrName=3&RFPPath=92), January 4, 2020.

**Table 2** Experimental variables

Variables		Categories
Gender	1	Female
	2	Male
Unemployment	1	No unemployment
	2	One year unemployed after graduation
	3	One year unemployed at time of application
Migrant background	1	Luxembourgish nationality & living in Luxembourg (referred to as: <i>Luxembourgish native</i> )
	2	Portuguese nationality & living in Luxembourg (referred to as: <i>Portuguese foreigner</i> )
	3	Luxembourgish-Portuguese nationality & living in Luxembourg (referred to as: <i>Luxembourgish-Portuguese foreigner</i> )
	4	French nationality & living in France (referred to as: <i>French border worker</i> )
	5	French nationality & living in Luxembourg (referred to as: referred to as: <i>French foreigner</i> )
	6	German nationality & living in Germany (referred to as: <i>German border worker</i> )

*excellent*).<sup>13</sup> The vignettes presented to the respondents were identical in both versions of the FSE. Figure 1 shows an example of a vignette presenting a male foreigner with Portuguese nationality and one year of unemployment after graduation applying for a catering job. The layout of the vignettes resembles the structure of a CV.

In the introduction to the rating task, the respondents in both FSE versions were instructed to assume that all applicants fulfilled the minimum language requirements as well as minimum requirements regarding educational credentials.<sup>14</sup> All respondents were informed that they would have the possibility to indicate selection criteria that are important for the respective position (and are missing in the vignettes) after the rating task.

In the HV version of the FSE, respondents were asked before the experiment whether they typically hired workers for that job type. Respondents who did not were excluded from the survey. The job types consisted of general occupational categories based on codes of the International Standard Classification of Occupations 2008 (ISCO-08; e.g., waiter/waitress, system administrator) that matched the sampled real vacancies for the RV version of the FSE (see Sect. 4.2). The rating task in the HV design was based on these generic job types as opposed to the RV design, where the rating task was based on the sampled real vacancies. We did not provide the respondents with details about the hypothetical job, as the main objective of this study is to test whether using real vacancies triggers more valid ratings of vignettes than using vague hypothetical vacancies does.

<sup>13</sup> This was the first of a series of questions that respondents were required to answer for each vignette. However, we focus on the hiring intentions of recruiters in this contribution.

<sup>14</sup> Luxembourg is a multilingual country, and job applicants are often required to speak several languages. It would have been difficult to create realistic applicant profiles if language had been included as a variable in addition to nationality. Nevertheless, these instructions might attenuate any statistical discrimination effect of migration background that is related to language and educational credentials (in both FSE designs).



Fig. 1 Example vignette

## 4.2 Sampling of real vacancies

We sampled job advertisements from various online job portals and company websites in Luxembourg based on pre-defined ISCO-08 categories. We focused on entry-level jobs located in Luxembourg. We searched seven generic job portals as well as occupation-specific job portals and company websites using a set of keywords and search categories that fit the ISCO categories. The online Appendix includes a complete list of these job portals and websites as well as the ISCO categories. The vacancies were collected manually; for the field of IT, additional vacancies were collected with the help of a web-scraping tool to facilitate the sampling process. Screenshots of all job advertisements were taken and later displayed to the respondents in the RV design before the FSE to remind them of the respective vacancy. When a job advertisement provided no contact information for recruiters, we called the company



and requested the name and email address of the person responsible for filling the vacancy. The vacancies were collected over 16 weeks between July and November 2018 (four weeks were reserved for telephone calls). We received email addresses of 203 recruiters in catering and 168 recruiters in IT.

### 4.3 Sampling strategy: hypothetical vacancies

Regarding the HV design of the FSE, we sampled respondents from publicly available lists of companies for each occupational field. Since the availability of such lists differed between occupations, we used different sources (e.g., yellow pages) and strategies for each occupation.<sup>15</sup> A detailed description of this sampling process is provided in the online Appendix. Company names, addresses, and phone numbers were collected manually and with the help of a computer program. If available, we collected the names and email addresses of the person responsible for recruitment in the respective companies or persons that were most likely involved in or had experiences in hiring within the respective field (e.g., managing directors or business owners). We called the collected companies and requested the contact information of a person actively involved in recruitment decisions. In total, we managed to collect the email addresses of 173 people in catering and 279 in IT.

### 4.4 Data collection and realized sample

We invited respondents via email to participate in an online survey about recruitment and operational staffing needs.<sup>16</sup> The survey was offered in three languages (French, German, and English). The data collection period started on November 22, 2018 and ended on January 25, 2019. In total, we sent out 823 invitations across the two occupations, and 140 respondents participated in the online survey (overall response rate of about 17%). The response rate for the RV version of the FSE was almost two times higher (22%) than the response rate for the HV version of the FSE (13%). The response rate was 19.7% in catering and 14.8% in IT. As in the whole sample, the response rate in the respective RV sample was considerably higher than in the HV sample. These results suggest that using real vacancies might actually increase interest in the survey and the willingness to participate.<sup>17</sup> The response rates are within the range of similar recruiter surveys using FSEs (e.g., Van Belle et al. 2018; Damelang et al. 2019).

Our sample consisted of 553 vignette ratings from 93 respondents who participated in the FSE; 300 observations from 50 respondents in catering and 253 observations from 43 respondents in IT.<sup>18</sup> We found differences in the distribution of key respondent characteristics between the two FSE versions (RV versus HV) in each occupation, particularly regarding

<sup>15</sup> Each company in Luxembourg is required to announce vacancies at Luxembourg's National Employment Agency (ADEM), which keeps records of company names and contact persons in each company. Unfortunately, we were not granted access to this list.

<sup>16</sup> M.I.S. Trend, a Swiss-based research institute, carried out the data collection on our behalf. The invitation requested that the email be forwarded to the person responsible for recruitment when the contacted person was not involved in their company's recruitment process (HV design) or not responsible for filling the given vacancy (RV design). As it is usually the case in self-administered surveys, this is no guarantee that the target person is the one who filled out our questionnaire.

<sup>17</sup> We reached similar response rates in our pilot study.

<sup>18</sup> We excluded respondents who dropped out of the survey before or during the FSE, but included respondents who dropped out after the FSE was completed to increase our overall sample size.

gender and citizenship. Therefore, we used entropy balancing (Hainmueller and Xu 2013) to adjust the distribution of gender, age, citizenship, and education in the two sub-samples (RV and HV) to the distribution of these characteristics in the overall sample in each occupation. We used these weights in our regression analyses.<sup>19</sup> Due to missing values, our final analytic sample was slightly reduced, to 282 observations from 47 recruiters in catering and 240 observations from 40 recruiters in IT. Tables 3 (catering) and 4 (IT) in the Appendix show that the weights account well for the differences in the distribution of respondent characteristics between the two FSE versions.

On average, each vignette has been evaluated 8.7 times in the RV version of the FSE (catering: 5.2, IT: 3.5) and 6.3 times in the HV version (catering: 3.2, IT: 3.7). In both FSE versions, respondents used the whole answer scale, and the distribution of vignette ratings was slightly left skewed in both occupations (see Figs. 4 and 5 in the Appendix). Given that the respondents were instructed to assume that requirements regarding educational qualifications are met, it is not surprising that the distribution tended toward positive values. Tables 5 and 6 in the Appendix further show that correlations between vignette variables were close to zero and not statistically significant for both occupations.

Because most of the vignettes presented potentially negative signals (e.g., unemployment) and because, according to our theory, real vacancies might help reduce the risk of response bias, the average vignette ratings were likely to be lower in the RV version of the FSE. However, a Wilcoxon-Mann-Whitney test to compare the average vignette ratings between the FSE versions showed no significant differences in both occupations.

## 4.5 Method

We estimated the effect of applicant characteristics on vignette ratings (i.e., recruiters' hiring intentions) using linear multilevel models. Given that each respondent rated six vignettes, the ratings were nested within respondents. Multilevel modeling accounted for this clustering in our data (Hox 2010). We estimated separate models for each occupation to account for possible general differences in staffing needs and recruitment processes that might affect our results. Moreover, the conducted FSE for each occupational field (catering, IT) can be seen as two separate experiments.

Our hypotheses pertain to the potential difference in the overall effects of migration background, gender, and unemployment based on the FSE design. We tested these hypotheses by combining categories 2 and 3 of unemployment and categories 2 to 6 of migrant background (see Table 2) into one category indicating unemployment experience and a foreign background, respectively. For each occupation, we estimated a model interacting the dummy for gender (1=female), unemployment (=1), and foreign background (=1) with the indicator for sample type (i.e., the FSE version). We further controlled for possible fatigue or primacy effects by including dummies for each vignette position (first to sixth). Our regression analyses were weighted (see Sect. 4.4). We calculated the marginal effects (Williams 2012) of each vignette variable on hiring intentions. Full models are shown in Table 7 in the Appendix.<sup>20</sup>

<sup>19</sup> We refer to the implications of applying weights in our regression analysis in our discussion (see Sect. 6).

<sup>20</sup> The syntax files used for analyses are available as supplementary material.

## 5 Results

### 5.1 Main results

The marginal effects of applicant characteristics on recruiters' hiring intentions are displayed in Fig. 2a (catering), b (IT). Having a foreign background significantly reduced recruiters' hiring intentions in the catering field (see Fig. 2a) at the 5% level in the HV version and at the 10% level in the RV version, matching the respective benchmark. We observed a slightly more negative effect of being a foreigner in the HV version (against our expectations), but the difference is not statistically significant. Hence, our findings contradict *Hypothesis 1* for migrant background in catering. Regarding IT jobs (Fig. 2b), we found no effect of migrant background in either FSE version, contradicting the benchmark; while the effects tended to be negative in the RV version and positive in the HV version, the difference was not statistically significant. Consequently, we again found no support for *Hypothesis 1* regarding migrant background in IT.

As for the applicants' gender, we observed positive effects of being female in catering in both FSE versions in line with the associated benchmark (Fig. 2a), but the effect was significant only in the RV version ( $p < 0.05$ ). This finding might be explained by the relatively high share of female workers in the catering sector in Luxembourg.<sup>21</sup> The observed effect of applicant gender was slightly less positive in the RV design, but the difference was not significant (contradicting *Hypothesis 1*). In turn, the gender effect on vignette ratings was close to zero and not significant in each IT FSE version (Fig. 2b). Unsurprisingly, we also found no support for *Hypothesis 1* regarding gender in IT.

Finally, as shown in Fig. 2a, unemployment significantly reduced recruiters' hiring intentions when applying for catering jobs in both FSE versions (RV:  $p < 0.01$ , HV:  $p < 0.01$ ). These findings support our benchmark regarding the effect of unemployment. Similar to the results regarding migrant background, we observed a slightly more negative effect of unemployment in the HV version; however, the difference was not significant (contradicting *Hypothesis 1* regarding unemployment in catering). Fig. 2b reveals that unemployment also had a highly significant negative effect on recruiters' hiring intentions in both IT FSE versions ( $p < 0.001$ ). The unemployment effect did not differ between the two FSE versions, and we found no support for *Hypothesis 1* in IT.

Our results also do not support *Hypothesis 2*. The differences in the effect sizes between the two FSE designs for unemployment do not statistically differ from those for gender or migration background. This holds for both occupations.

### 5.2 Additional analyses

We repeated our analyses using the operationalization of vignette variables presented in Table 1 to test whether we find similar effects and design differences when considering specific categories of migration background and unemployment. Figure 3 shows the marginal effects resulting from these models in catering (Fig. 3a) and IT jobs (Fig. 3b). The full models are presented in Table 8 in the Appendix.

As shown in Fig. 3a for jobs in catering, the vignette ratings were on average lower toward German border workers (HV:  $p < 0.01$ , RV:  $p < 0.05$ ), French foreigners (HV:  $p < 0.05$ , RV:  $p = 0.058$ ), and French border workers (HV:  $p = 0.065$ , RV:  $p < 0.05$ ) compared

<sup>21</sup> About 60% of the waitstaff is female based on data from the Structure of Earnings Survey 2014. The information was provided to us by STATEC, Luxembourg's statistical office, upon request.

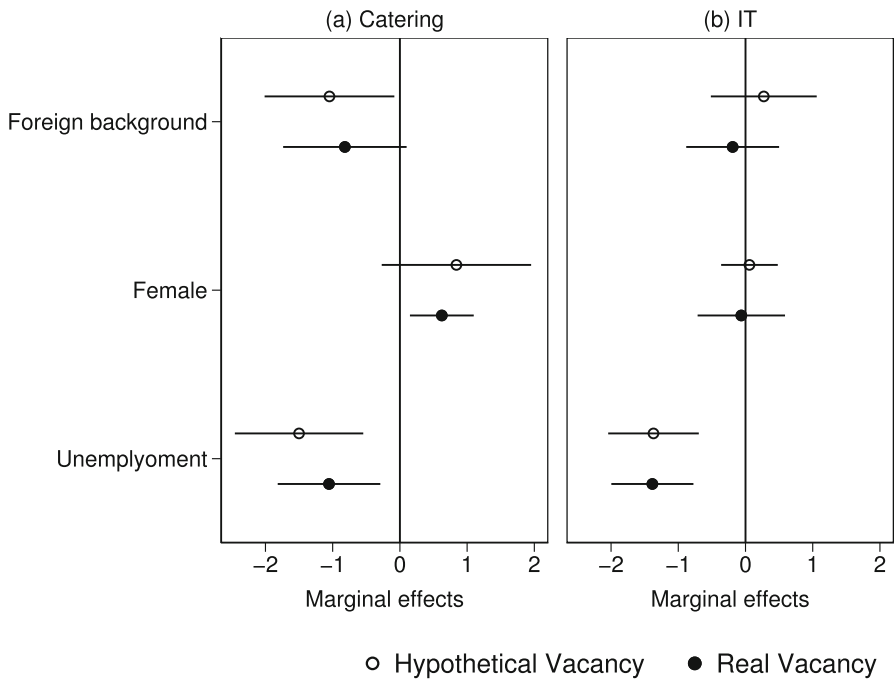
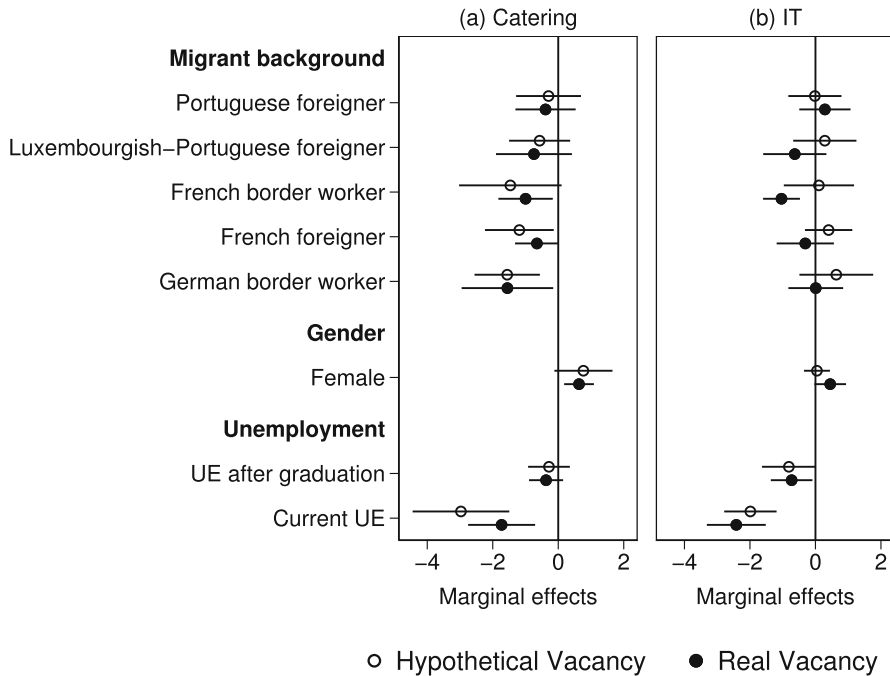


Fig. 2 Marginal effects of applicant characteristics based on full model presented in Table 7

to Luxembourgish natives (HV:  $p < 0.01$ , RV:  $p < 0.05$ ). For both categories indicating French nationality, the effects seemed to be more negative in the HV than the RV design. However, we did not find significant differences in the effects of the above characteristics between the FSE versions. This is in line with our findings regarding *Hypothesis 1* when looking at the combined effect of migrant background. Figure 3b shows that there were no significant effects in the HV version for any category of migrant background in IT. Also, whereas in the RV version most effects tended to be negative, here all effects were positive, contrary to our benchmarks. We only observed a significant effect for French border workers in the RV version ( $p < 0.001$ ). Most importantly, the difference in this effect between the two FSE versions was significant at the 10% level ( $p = 0.064$ ), supporting *Hypothesis 1*.

Our additional analyses revealed no substantial changes in the effect of applicants' gender on vignette ratings.

Regarding the negative unemployment effect in catering (Fig. 3a), we found that only current unemployment significantly reduced recruiters' hiring intentions for both FSE versions ( $p < 0.001$ ). The effect was slightly less negative in the RV version, but the difference was not statistically significant. Regarding IT jobs (Fig. 3b), current unemployment had a highly significant negative effect on vignette ratings in both FSE versions ( $p < 0.001$ ); however, the effect seemed to be more negative in the RV version. The effect of past unemployment was (marginally) significant at the 5% level in both FSE designs (HV:  $p = 0.052$ , RV:  $p < 0.05$ ). None of the differences in unemployment effects between the FSE versions were significant. Consequently, in line with our results reported in Sect. 5.1, we found no support for *Hypothesis 1* regarding unemployment in either catering or IT. Finally, we found no support for *Hypothesis 2* in either occupation.



**Fig. 3** Marginal effects of applicant characteristics by single categories based on full model presented in Table 8

## 6 Discussion and conclusions

In this study, we conducted an FSE for two occupational fields (catering and IT) in Luxembourg to examine whether the effects of applicant characteristics on observed hiring intentions differ when the vignette ratings refer to a real instead of a hypothetical vacancy. We focused on three applicant characteristics in our analyses: migration background, gender, and unemployment.

Estimating multilevel regressions, our results regarding the main effects of applicant characteristics largely correspond to our benchmarks. However, we observed some differences based on the occupational field. Regarding the migration background, applicants with a foreign background face disadvantages in catering, which holds for both FSE versions. In the RV version, the effect was slightly less negative than in the HV version. In contrast, migration background had no significant effect on vignette ratings in IT, although the respective effect tended toward different directions depending on the FSE design. The effect of gender (1=female) tended to be positive in both FSE versions in catering but only in the HV version in IT. In the RV version in IT, the effect of gender was negative. Women might have an advantage in catering because service jobs are typically considered to be a “female” job (e.g., Booth and Leigh 2010). A female disadvantage in IT is, in turn, in line with prior research suggesting that women are discriminated in STEM (science, technology, engineering, math) occupations (e.g., Kübler et al. 2018). However, the effect of gender was not significant in IT in our study. Finally, we found very similar results regarding unemployment for both occupations: in both cases, unemployment significantly reduced hiring intentions. The respective

effect rarely differed between the FSE designs in IT but was slightly less negative in the RV version in catering. The differing patterns in the observed effects between occupations might be explained by general structural differences between the two sectors.

Overall, none of the observed differences in the effects between the FSE versions were statistically significant. A more fine-grained analysis of single categories of migration background revealed that French border workers were particularly disadvantaged in IT in the RV version. In the HV version in IT, the effect of being a French border worker tended to be positive but was not significant. The observed difference was statistically significant at the 10% level. This finding lends some support to the use of real-world vacancies, as our benchmark suggested a negative effect of being a non-native person on recruiters' hiring intentions. Excluding this finding, our results provide little evidence for differences in the effect of applicant characteristics on vignette ratings between RV and HV FSE designs.

We observed in this study and the pilot study that response rates in the RV design were about twice those in the HV design of the FSE. If response rates are of special concern, efficient sampling of real vacancies might be an attractive option. Most importantly, higher involvement of recruiters in the FSE due to real vacancies probably also contributes to higher quality of answers and higher internal validity of the results.

This study has some limitations, which make an overall conclusion about the impact of the FSE design on hiring intentions difficult. Our findings must therefore be interpreted with caution. First, our small sample of real vacancies made it impossible to employ a true split ballot experiment. A random assignment of respondents to one of the two FSE designs would have provided a more robust estimate of the effect of using real vacancies. Instead, we compared two different samples that slightly differed in the composition of the respondents: (i) actual recruiters responsible for filling a real vacancy in their company and (ii) a mix of human resource professionals, managers, and directors. Hence, the comparability of the two samples is limited. In our regression analyses, we have used weights to balance both FSE samples for key respondent characteristics. However, using weights does not change our main findings compared to an analysis without weights. Additional analyses, which we present in the online Appendix, further revealed that the randomization processes within our experiment were successful. We found correlations very close to zero and not statistically significant between vignette variables and observed respondent characteristics (see Table A3 and A4). We also estimated fixed effects regressions with vignette variables as predictors to test the influence of unobserved respondent characteristics (Auspurg and Hinz 2015). The correlation ( $u_j, X$ ) between the error term and control variables can be taken as indicator for omitted respondent characteristics that influence vignette ratings. In both occupations, these correlations were close to zero (see Table A5). Second, to reliably test whether using real vacancies in FSEs leads to more valid vignette ratings than using hypothetical vacancies, behavioral benchmarks reflecting the true effects of our applicant characteristics on hiring decisions would have been desirable; however, these were not available for Luxembourg. Hence, we compared our results to "theoretical" benchmarks derived from previous experimental research from other contexts (see Sect. 3). Third, our sample size was rather small. Some of our null findings might be due to limited statistical power because of the low number of observations.

Moreover, we were unable to go beyond a broad comparison of two occupations (catering and IT specialists) and consider potential contextual moderators more in detail. The job requirements as well as structural characteristics (e.g., unemployment rates and labor market tightness) typically differ between labor market sectors. These factors may have an impact on how employers interpret and act upon the information provided in vignettes. While this limits the generalizability of our findings concerning the main effects of applicant characteristics on recruiters' hiring intentions, it should not affect the generalizability of our methodological

effects. Nevertheless, as stated in Sect. 3, differences in vignette ratings when real vacancies as opposed to hypothetical ones are used might be particularly pronounced if the urgency of finding suitable candidates is high. This might be the case, for example, if recruitment is perceived to be difficult. Most respondents in our sample generally found it difficult to recruit suitable workers,<sup>22</sup> which is why a comparative analysis concerning the perceived difficulty to find applicants was not possible due to low variance.

Finally, the idiosyncrasies of Luxembourg limit the comparability of our results with other studies. For example, the Luxembourgish labor market is unique in the multi-cultural composition of its workforce. As aforementioned, different outcomes particularly regarding the main effects of applicant characteristics on recruiters' hiring intentions may be expected in other contexts.

The question of whether we should use real or hypothetical vacancies when studying hiring prompts more general considerations about sampling and inference. Researchers will have to assess whether they wish to make statements about mechanisms at the level of companies, occupations, vacancies, or recruiters or a combination of these. For example, if the main interest is to investigate potential discrimination, it will be beneficial to sample randomly from a pool of actually available vacancies, as they may differ systematically from filled positions in a given occupation. If the primary aim is to advance our general understanding of recruiter behavior in a given sector, on the other hand, presenting hypothetical vacancies to a random selection of recruiters from this sector might be adequate.

This study draws attention to potential issues surrounding the routine use of hypothetical vacancies in FSEs, an area that has been widely neglected in previous studies on recruiters' hiring intentions. Our results suggest that using hypothetical vacancies might be, within the limits of FSEs, a valid approach. Note, however, that the effect of using real vacancies might have been more pronounced if the differences between our two samples would have been larger (e.g., recruiters with only general recruiting experience vs. recruiters with job-specific recruiting knowledge). However, identifying the effect of using real vacancies would have been more difficult with such an approach, as both the type of sample and the type of vacancy potentially affect the perceived realism of the rating task.

In any case, the potential effect of using real-world vacancies in FSEs on response behavior deserves further scrutiny. This seems particularly important given the potential methodological and practical implications for the increasing number of FSEs in employer studies. Substantial and theoretically plausible differences between FSE designs would further encourage the use of real vacancies as the internal and external validity is likely increased. Besides aspects that could not be analyzed in this study (e.g., the role of recruitment difficulties), the present analyses could be extended comparing the survey responses from vignette ratings referring to real vacancies with behavioral data to test the internal and external validity of the results obtained from both types of FSE (i.e., RV versus HV design). In doing so, researchers should ensure comparability in experimental designs and sample composition between studies (Petzold and Wolbring 2019).

A better understanding of the implications of the differences in FSE designs between studies is surely needed to establish best practices in employer studies. Prior research on general methodological questions related to the design of FSEs (e.g., Auspurg et al. 2019; Auspurg and Jäckle 2017; Sauer et al. 2011; Shamon et al. 2019) continues to be highly relevant for the application of FSEs to study employer preferences, as it provides common guidelines for designing vignettes and answer scales. More studies on methodological aspects

<sup>22</sup> Respondents were asked about the difficulty of filling the advertised position (RV sample) or the selected profession (HV sample) on a scale from 1=*Very easy* to 4=*Very difficult*.



of FSEs are required, however, to advance our understanding of their possibilities and limits in research on hiring. We hope that our comparison of designs using real and hypothetical vacancies contributes to this emerging strand of methodological inquiry.

**Funding Information** The study was funded by an Internal Research Project Grant of the University of Luxembourg [IRP17 - EDYPOLU].

**Availability of data and material** The data will be made available in spring 2021 as scientific use files at <https://osf.io/f6yba/>.

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflicts of interest.

**Ethics approval** The study received ethics approval by the Ethics Review Panel of the University of Luxembourg [ERP 18-009].

**Consent to participate** The study participants provided informed consent.

**Consent for publication** The study participants were informed that the results would be published in scientific journals.

**Code availability** The authors provide the syntax files used for analyses as supplementary material to this article.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix

**Table 3** Distribution of respondent characteristics across FSE designs (catering)

Variable	Overall	HV design		RV design	
		Unweighted	Weighted	Unweighted	Weighted
Age ( $\emptyset$ )	46	46	46	45	46
Female (%)	47	26	47	61	47
Luxembourgish citizenship (%)	47	58	47	39	47
<i>Highest educational certificate (%)</i>					
Compulsory school	8	11	8	7	8
Vocational training	27	37	27	21	27
University entry qualification	15	11	15	17	15
University degree	38	37	38	38	38
Other	13	5	12	17	12

$N = 47$  recruiters overall; RV:  $N = 28$  recruiters, HV:  $N = 19$  recruiters

Italic: headings for categorical variables with 3 or more categories.

**Table 4** Distribution of respondent characteristics across FSE designs (IT)

Variable	Overall		HV design		RV design	
			Unweighted	Weighted	Unweighted	Weighted
Age ( $\emptyset$ )	43		47	43	37	43
Female (%)	33		27	33	39	33
Luxembourgish citizenship (%)	25		23	25	28	25
<i>Highest educational certificate (%)</i>						
Compulsory school	0		0	0	0	0
Vocational training	8		5	7	11	15
University entry qualification	13		18	13	6	12
University degree	73		64	72	83	73
Other	8		14	8	0	0

*N* = 40 recruiters overall; RV: *N* = 18 recruiters, HV: *N* = 22 recruiters

*Italic*: headings for categorical variables with 3 or more categories.

**Table 5** Correlations between vignette variables: catering

	1	2	3	4	5	6	7	8	9	10
1 Female	1.000									
2 No unemployment	0.030	1.000								
3 Unemployed after graduation	0.000	-0.500*	1.000							
4 Unemployed at time of application	-0.030	-0.500*	-0.500*	1.000						
5 Luxembourgish native	-0.010	0.027	-0.013	-0.013	1.000					
6 Portuguese foreigner	0.010	0.007	0.007	-0.013	-0.200*	1.000				
7 Luxembourgish-Portuguese foreigner	-0.010	0.007	0.007	-0.013	-0.200*	-0.200*	1.000			
8 French border worker	-0.029	-0.013	0.007	0.007	-0.200*	-0.200*	-0.200*	1.000		
9 French foreigner	0.010	-0.013	-0.013	0.027	-0.200*	-0.200*	-0.200*	-0.200*	1.000	
10 German border worker	0.029	-0.013	0.007	0.007	-0.200*	-0.200*	-0.200*	-0.200*	-0.200*	1.000

Pairwise correlations between all levels of vignette variables (as dummies), Pearson's correlation coefficient.

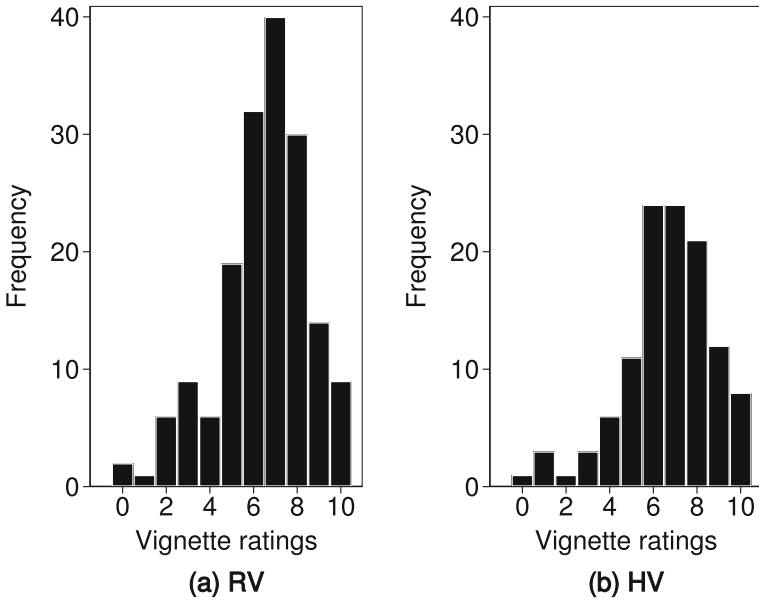
\*  $p < 0.05$

**Table 6** Correlations between vignette variables: IT

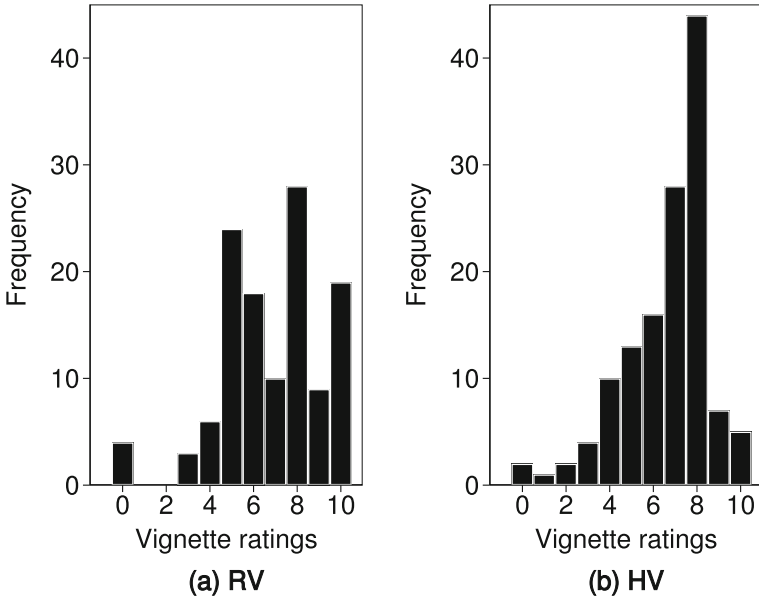
	1	2	3	4	5	6	7	8	9	10
1 Female	1.000									
2 No unemployment	-0.053	1.000								
3 Unemployed after graduation	0.000	-0.500*	1.000							
4 Unemployed at time of application	0.053	-0.500*	-0.500*	1.000						
5 Luxembourgish native	-0.022	-0.008	-0.079	0.087	1.000					
6 Portuguese foreigner	-0.067	0.040	0.040	-0.079	-0.200*	1.000				
7 Luxembourgish-Portuguese foreigner	0.045	0.016	0.040	-0.055	-0.200*	-0.200*	1.000			
8 French border worker	0.022	-0.079	0.040	0.040	-0.200*	-0.200*	-0.200*	1.000		
9 French foreigner	0.022	-0.055	0.040	0.016	-0.200*	-0.200*	-0.200*	-0.200*	1.000	
10 German border worker	0.000	0.087	-0.079	-0.008	-0.200*	-0.200*	-0.200*	-0.200*	-0.200*	1.000

Pairwise correlations between all levels of vignette variables (as dummies), Pearson's correlation coefficient.

\*  $p < 0.05$



**Fig. 4** Distribution of vignette ratings in the catering sector. RV:  $N = 168$  vignette ratings from 28 recruiters; HV:  $N = 114$  vignette ratings from 19 recruiters



**Fig. 5** Distribution of vignette ratings in the IT sector. RV:  $N = 108$  vignette ratings from 18 recruiters; HV:  $N = 132$  vignette ratings from 22 recruiters

**Table 7** Results from multilevel regressions predicting hiring intentions; full models

	Catering	IT
Foreign background (Ref.: Luxembourgish)	-1.048* (0.492)	0.272 (0.402)
Real vacancy (Ref.: Hypothetical vacancy)	-0.059 (0.957)	1.198* (0.581)
Foreign background × Real vacancy	0.229 (0.680)	-0.463 (0.534)
Female (Ref.: Male)	0.841 (0.567)	0.059 (0.215)
Female × Real vacancy	-0.217 (0.617)	-0.121 (0.395)
Unemployment (Ref.: No unemployment)	-1.500** (0.487)	-1.369*** (0.344)
Unemployment × Real vacancy	0.446 (0.624)	-0.016 (0.464)
<i>Vignette position</i> (Ref.: Position 1)		
Position 2	0.727 (0.725)	0.802 (0.528)
Position 3	1.368 <sup>+</sup> (0.738)	0.396 (0.491)
Position 4	1.100 (1.154)	0.412 (0.537)
Position 5	1.008 (1.057)	0.844 (0.681)
Position 6	1.061 <sup>+</sup> (0.608)	1.090 <sup>+</sup> (0.593)
<i>Vignette position × Real vacancy</i>		
Position 2 × Real vacancy	-0.029 (0.880)	-0.970 (0.840)
Position 3 × Real vacancy	-0.523 (0.928)	-0.293 (0.609)
Position 4 × Real vacancy	-0.526 (1.197)	-0.103 (0.696)
Position 5 × Real vacancy	-1.129 (1.283)	0.019 (0.849)
Position 6 × Real vacancy	-0.924 (0.725)	-0.473 (0.811)
Constant	7.139*** (0.877)	6.798*** (0.483)

**Table 7** continued

	Catering	IT
Variance: Recruiter	0.964 (0.372)	1.484 (0.803)
Variance: Vignettes	3.064*** (0.599)	2.136*** (0.441)

Catering:  $N = 282$  observations from 47 recruiters; IT:  $N = 240$  observations from 40 recruiters. Weighted data. Standard errors in parentheses.

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , +  $p < 0.10$

Italic: headings for variables and interactions with 3 or more categories.

**Table 8** Results from multilevel regressions predicting hiring intentions; full models (single categories of vignette variables)

	Catering		IT	
<i>Migrant background</i> (Ref.: Luxembourgish)				
Portuguese foreigner	-0.296	(0.505)	-0.017	(0.414)
Luxembourgish-Portuguese foreigner	-0.568	(0.476)	0.286	(0.493)
French border worker	-1.462 <sup>+</sup>	(0.798)	0.106	(0.547)
French foreigner	-1.187*	(0.536)	0.402	(0.369)
German border worker	-1.558**	(0.509)	0.637	(0.575)
Real vacancy (Ref.: Hypothetical vacancy)	-0.128	(0.833)	0.922 <sup>+</sup>	(0.541)
<i>Migrant background × Real vacancy</i>				
Portuguese foreigner × Real vacancy	-0.092	(0.689)	0.306	(0.575)
Luxembourgish-Portuguese foreigner × Real vacancy	-0.173	(0.759)	-0.917	(0.696)
French border worker × Real vacancy	0.465	(0.903)	-1.143 <sup>+</sup>	(0.618)
French foreigner × Real vacancy	0.536	(0.635)	-0.713	(0.579)
German border worker × Real vacancy	0.006	(0.876)	-0.627	(0.715)
<i>Female</i> (Ref.: Male)				
Female × Real vacancy	-0.134	(0.508)	0.404	(0.319)
<i>Unemployment</i> (Ref.: No unemployment)				
Unemployment after graduation	-0.283	(0.325)	-0.812 <sup>+</sup>	(0.419)
Currently unemployed	-2.971***	(0.753)	-1.987***	(0.408)
<i>Unemployment × Real vacancy</i>				
Unemployment after graduation × Real vacancy	-0.089	(0.420)	0.085	(0.529)
Currently unemployed × Real vacancy	1.242	(0.916)	-0.430	(0.614)
<i>Vignette position</i> (Ref.: Position 1)				
Position 2	0.518	(0.567)	1.120**	(0.433)
Position 3	0.447	(0.514)	0.224	(0.345)
Position 4	0.968	(0.818)	0.349	(0.528)
Position 5	0.318	(0.645)	0.957	(0.719)
Position 6	0.400	(0.461)	0.843	(0.551)
<i>Vignette position × Real vacancy</i>				
Position 2 × Real vacancy	-0.376	(0.744)	-0.882	(0.660)
Position 3 × Real vacancy	-0.360	(0.699)	-0.039	(0.444)



**Table 8** continued

	Catering		IT	
Position 4 × Real vacancy	−0.798	(0.906)	0.345	(0.653)
Position 5 × Real vacancy	−0.759	(0.950)	0.490	(0.832)
Position 6 × Real vacancy	−0.794	(0.598)	−0.250	(0.706)
Constant	7.668***	(0.732)	6.825***	(0.459)
Variance: Recruiter	1.106	(0.361)	1.557	(0.452)
Variance: Vignettes	2.210***	(0.361)	1.698*	(0.405)

Catering:  $N = 282$  observations from 47 recruiters; IT:  $N = 240$  observations from 40 recruiters. Weighted data. Operationalization of vignette variables as shown in Table 2. Standard errors in parentheses.

\*\*\*  $p < 0.001$ , \*\*  $p < 0.01$ , \*  $p < 0.05$ , +  $p < 0.10$

Italic: headings for variables and interactions with 3 or more categories.

## References

- Ajzen, I., Brown, T.C., Carvajal, F.: Explaining the discrepancy between intentions and actions: the case of hypothetical bias in contingent valuation. *Personal Soc. Psychol. Bull.* **30**(9), 1108–1121 (2004)
- Atzmüller, C., Steiner, P.M.: Experimental vignette studies in survey research. *Methodol. Eur. J. Res. Methods Behav. Soc. Sci.* **6**(3), 128–138 (2010)
- Auer, D., Bonoli, G., Fossati, F., Liechti, F.: The matching hierarchies model: evidence from a survey experiment on employers' hiring intent of immigrant applicants. *Int. Migr. Rev.* **53**(1), 90–121 (2019)
- Auspurg, K., Hinz, T.: *Factorial Survey Experiments*. Sage Publications, Los Angeles (2015)
- Auspurg, K., Jäckle, A.: First equals most important? order effects in vignette-based measurement. *Sociol. Methods Res.* **46**(3), 490–539 (2017)
- Auspurg, K., Hinz, T., Liebig, S., Sauer, C.: The Factorial Survey as a Method for Measuring Sensitive Issues. In: Engel, U., Jann, B., Lynn, P., Scherpenzeel, A., Sturgis, P. (eds.) *Improving Survey Methods: Lessons from Recent Research*, pp. 137–149. Routledge, New York (2015)
- Auspurg, K., Hinz, T., Walzenbach, S.: Are factorial survey experiments prone to survey mode effects? In: Lavrakas PJ, Traugott MW, Kennedy C, Holbrook AL, de Leeuw ED, West BT (eds) *Experimental Methods in Survey Research: Techniques that Combine Random Sampling with Random Assignment*. Wiley, Hoboken, pp. 371–392 (2019)
- Baert, S., (2018) Hiring discrimination: an overview of (Almost) all correspondence experiments since. In: Gaddis, S.M. (ed.) *Audit Studies: Behind the Scenes with Theory, Method, and Nuance*, method, ser edn, pp. 63–77. Springer, New York (2005)
- Baert, S., As, De Pauw: Is ethnic discrimination due to distaste or statistics? *Econ. Lett.* **125**(2), 270–273 (2014)
- Baert, S., Cockx, B., Gheyle, N., Vandamme, C.: Is there less discrimination in occupations where recruitment is difficult? *Ind. Labor Relations Rev.* **68**(3), 467–500 (2015)
- Bertogg, A., Imdorf, C., Hyggen, C., Parsanoglou, D., Stoilova, R.: Gender discrimination in the hiring for skilled professionals in two male-dominated occupational fields—a factorial survey experiment with real-world vacancies and recruiters in four European countries. *Kolner. Z. Soz. Sozpsychol.* (2020). <https://doi.org/10.1007/s11577-020-00671-6>
- Biesma, R.G., Pavlova, M., van Merode, G.G., Groot, W.: Using conjoint analysis to estimate employers preferences for key competencies of master level Dutch graduates entering the public health field. *Econ. Educ. Rev.* **26**(3), 375–386 (2007)
- Bills, D.B., Stasio, V.D., Gërkhani, K.: The demand side of hiring: employers in the labor market. *Ann. Rev. Sociol.* **43**, 6.1–6.20 (2017)
- Birkelund, G.E., Janz, A., Larsen, E.N.: Do males experience hiring discrimination in female-dominated occupations? An overview of field experiments since 1996. GEMM working paper, <https://gemm2020.eu/wp-content/uploads/2018/12/Gender-Discrimination-a-summary-1.pdf> (2019)
- Booth, A., Leigh, A.: Do employers discriminate by gender? a field experiment in female-dominated occupations. *Econ. Lett.* **107**, 236–238 (2010)
- Brewer, M.B., Crano, W.D.: *Research Design and Issues of Validity*. In: Reis, H.T., Judd, C.M. (eds.) *Handbook of research methods in social and personality psychology*, 2nd edn, pp. 11–26. Cambridge University Press, New York (2014)

- Carlsson, M.: Does hiring discrimination cause gender segregation in the Swedish labor market? *Fem. Econ.* **17**(3), 71–102 (2011)
- Damelang, A., Abraham, M.: You can take some of it with you! a vignette study on the acceptance of foreign vocational certificates and ethnic inequality in the German labor market. *Z. Soziol.* **45**(2), 91–106 (2016)
- Damelang, A., Abraham, M., Ebensperger, S., Stumpf, F.: The hiring prospects of foreign-educated immigrants: a factorial survey among German employers. *Work Employ. Soc.* **33**, 1–20 (2019)
- Damelang, A., Ebensperger, S., Stumpf, F.: Foreign credential recognition and immigrants' chances of being hired for skilled jobs—evidence from a survey experiment among employers. *Soc. Forces* (2020). <https://doi.org/10.1093/sf/soz154>
- De Wolf, I., Van Der Velden, R.: Selection processes of three types of academic jobs. An experiment among Dutch employers of social sciences graduates. *Eur. Sociol. Rev.* **17**(3), 317–330 (2001)
- Di Stasio, V.: Education as a signal of trainability: results from a vignette study with Italian employers. *Eur. Sociol. Rev.* **30**(6), 796–809 (2014)
- Di Stasio, V.: Who is ahead in the labor queue? institutions' and employers' perspective on overeducation, undereducation, and horizontal mismatches. *Sociol. Educ.* **90**(2), 109–126 (2017)
- Di Stasio, V., Gërxhani, K.: Employers' social contacts and their hiring behavior in a factorial survey. *Soc. Sci. Res.* **53**, 93–107 (2015)
- Di Stasio, V., Van De Werfhorst, H.G.: Why does education matter to employers in different institutional contexts? a vignette study in England and the Netherlands. *Soc. Forces* **95**(1), 77–106 (2016)
- Dülmer, H.: Experimental plans in factorial surveys: random or quota design? *Sociol. Methods Res.* **35**(3), 382–409 (2007)
- Eriksson, S., Rooth, D.O.: Do employers use unemployment as a sorting criterion when hiring? evidence from a field experiment. *Am. Econ. Rev.* **104**(3), 1014–1039 (2014)
- Fernandez, R.M., Mors, M.L.: Competing for jobs: labor queues and gender sorting in the hiring process. *Soc. Sci. Res.* **37**(4), 1061–1080 (2008)
- Hainmueller, J., Xu, Y.: Ebalance: a stata package for entropy balancing. *J. Stat. Softw.* **54**(7), 1–18 (2013)
- Hainmueller, J., Hangartner, D., Yamamoto, T.: Validating vignette and conjoint survey experiments against real-world behavior. *Proc. Natl. Acad. Sci.* **112**(8), 2395–2400 (2015)
- Hox, J.J.: *Multilevel analysis: techniques and applications*, 2nd edn. Routledge, New York (2010)
- Humburg, M., Van Der Velden, R.: Skills and the graduate recruitment process: evidence from two discrete choice experiments. *Econ. Educ. Rev.* **49**, 24–41 (2015)
- Imdorf, C., Shi, L.P., Sacchi, S., Samuel, R., Hyggen, C., Stoilova, R., Yordanova, G., Boyadjieva, P., Ilieva-Trichkova, P., Parsanoglou, D., Yfanti, A.: Scars of early job insecurity across Europe: insights from a multi-country employer study, risk factors and policies. In: Hvinden, B., Hyggen, C., Schoyen, M.A., Sirovátka, T. (eds.) *Youth Unemployment and Job Insecurity in Europe*, pp. 93–116. Edward Elgar Publishing, Cheltenham (2019)
- Karpinska, K., Henkens, K., Schippers, J.: The recruitment of early retirees: a vignette study of the factors that affect managers' decisions. *Ageing Soc.* **31**(4), 570–589 (2011)
- Keuschnigg, M., Wolbring, T.: The use of field experiments to study mechanisms of discrimination. *Anal. Krit.* **38**(1), 179–202 (2016)
- Koch, A.J., D'Mello, S.D., Sackett, P.R.: A meta-analysis of gender stereotypes and bias in experimental simulations of employment decision making. *J. Appl. Psychol.* **100**(1), 128–161 (2015)
- Kroft, K., Lange, F., Notowidigdo, M.J.: Duration dependence and labor market conditions: evidence from a field experiment. *Q. J. Econ.* **128**, 1123–1167 (2013)
- Krosnick, J.A., Judd, C.M., Wittenbrink, B.: The measurement of attitudes. In: Albarracín, D., Johnson, B.T., Zanna, M.P. (eds.) *The Handbook of Attitudes*, pp. 21–78. Psychology Press, New York (2014)
- Kübler, D., Schmid, J., Stüber, R.: Gender discrimination in hiring across occupations: a nationally-representative vignette study. *Labour Econ.* **55**, 215–229 (2018)
- Kuhfeld, W.F.: *Experimental design: efficiency, coding, and choice designs*. Marketing Research Methods in SAS, pp. 53–241. SAS Institute, Cary (2010)
- Liechti, F., Fossati, F., Bonoli, G., Auer, D.: The signalling value of labour market programmes. *Eur. Sociol. Rev.* **33**(2), 257–274 (2017)
- McDonald, P.: How factorial surveys analysis improves our understanding of employer preferences. *Schweiz. Z. Soziol.* **45**(2), 237–260 (2019)
- Mergener, A., Maier, T.: Immigrants' chances of being hired at times of skill shortages: results from a factorial survey experiment among German employers. *J. Int. Migr. Integr.* pp 1–23 (2018)
- NEGOTIATE: NEGOTIATE Employer Survey. Scientific Use File. Data Documentation, OsloMet, Oslo (2020)

- Nunley, J.M., Pugh, A., Romero, N., Seals, R.A.: The effects of unemployment and underemployment on employment opportunities: results from a correspondence audit of the labor market for college graduates. *ILR Rev.* **70**(3), 642–670 (2017)
- Oberholzer-Gee, F.: Nonemployment stigma as rational herding: a field experiment. *J. Econ. Behav. Organ.* **65**(1), 30–40 (2008)
- Oesch, D., Lipps, O., McDonald, P.: The wage penalty for motherhood: evidence on discrimination from panel data and a survey experiment for Switzerland. *Demogr. Res.* **37**(1), 1793–1824 (2017)
- Pager, D., Quillian, L.: Walking the talk? what employers say versus what they do. *Am. Sociol. Rev.* **70**(3), 355–380 (2005)
- Pedulla, D.S.: Penalized or protected? gender and the consequences of nonstandard and mismatched employment histories. *Am. Sociol. Rev.* **81**(2), 262–289 (2016)
- Pedulla, D.S.: How race and unemployment shape labor market opportunities: additive, amplified, or muted effects? *Soc. Forces* **96**(March), 1477–1506 (2018)
- Petersen, T., Tostgad, T.: Getting the offer: sex discrimination in hiring. *Res. Soc. Stratif. Mobil.* **24**(3), 239–257 (2006)
- Petzold, K.: The role of international student mobility in hiring decisions. A vignette experiment among German employers. *J. Educ. Work* **30**(8), 893–911 (2017)
- Petzold, K., Wolbring, T.: What can we learn from factorial surveys about human behavior?: a validation study comparing field and survey experiments on discrimination. *Methodology* **15**(1), 19–28 (2019)
- Protsch, P., Solga, H.: How employers use signals of cognitive and noncognitive skills at labour market entry: insights from field experiments. *Eur. Sociol. Rev.* **31**(5), 521–532 (2014)
- Protsch, P., Solga, H.: Going across Europe for an apprenticeship? a factorial survey experiment on employers' hiring preferences in Germany. *J. Eur. Soc. Policy* **27**(4), 387–399 (2017)
- Quillian, L., Heath, A., Pager, D., Midtbøen, A.H., Fleischmann, F., Hexela, O.: Do some countries discriminate more than others? evidence from 97 field experiments of racial discrimination in hiring. *Sociol. Sci.* **6**, 467–496 (2019)
- Riach, P.A., Rich, J.: Deceptive field experiments of discrimination: are they ethical? *Kyklos* **57**(3), 457–470 (2004)
- Rossi, P.: Vignette analysis: uncovering the normative structure of complex judgments. In: Merton, R., Coleman, J., Rossi, P. (eds.) *Qualitative and Quantitative Social Research: Papers in Honor of Paul F. Lazarsfeld*, pp. 176–186. Free Press, New York (1979)
- Sauer, C., Hinz, T., Auspurg, K., Liebig, S., Hinz, T., Liebig, S.: The application of factorial surveys in general population samples: the effects of respondent age and education on response times and response consistency. *Surv. Res. Methods* **5**(3), 26 (2011)
- Schwarz, N.: Self-reports: how the questions shape the answers. *Am. Psychol.* **54**(2), 93–105 (1999)
- Shamon, H., Dülmer, H., Giza, A.: The factorial survey: the impact of the presentation format of vignettes on answer behavior and processing time. *Sociol. Methods Res.* (2019). <https://doi.org/10.1177/0049124119852382>
- Shi, L.P., Imdorf, C., Samuel, R., Sacchi, S.: How unemployment scarring affects skilled young workers: evidence from a factorial survey of Swiss recruiters. *J. Labour Mark. Res.* **52**(1), 7 (2018)
- Sniderman, P.M.: Some advances in the design of survey experiments. *Ann. Rev. Polit. Sci.* **21**(1), 259–275 (2018)
- Todesco, L.: The factorial survey in social research: its pros and cons. *Adv. Sociol. Res.* **23**, 1–29 (2017)
- Van Beek, K.W.H., Koopmans, C.C., van Praag, B.M.S.: Shopping at the labour market: a real tale of fiction. *Eur. Econ. Rev.* **41**, 295–317 (1997)
- Van Belle, E., Caers, R., De Couck, M., Di Stasio, V., Baert, S.: Why are employers put off by long spells of unemployment? *Eur. Sociol. Rev.* **34**(6), 1–17 (2018)
- Van Belle, E., Caers, R., De Couck, M., Di Stasio, V., Baert, S.: The signal of applying for a job under a vacancy referral scheme. *Ind. Relat. (Berkeley)* **58**(2), 251–274 (2019)
- Wallander, L.: 25 years of factorial surveys in sociology: a review. *Soc. Sci. Res.* **38**(3), 505–520 (2009)
- Williams, R.: Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata J.* **12**(2), 308–331 (2012)
- Wilson, A.: A silver lining for disadvantaged youth on the apprenticeship market—an experimental study of employers' hiring preferences. *J. Vocat. Educ. Train* (2019). <https://doi.org/10.1080/13636820.2019.1698644>
- Zschirnt, E., Ruedin, D.: Ethnic discrimination in hiring decisions: a meta-analysis of correspondence tests 1990–2015. *J. Ethn. Migr. Stud.* **42**(7), 1115–1134 (2016)