methods, data, analyses | Vol. 19(2), 2025, pp. 87-124 DOI: 10.12758/mda.2025.08

Researching the Moment of Truth: An Experiment Comparing In-the-Moment and Conventional Web Surveys to Investigate Online Job Applications

Carlos Ochoa

Research and Expertise Centre for Survey Methodology, Universitat Pompeu Fabra

Abstract

Understanding how people seek and apply for jobs online is crucial for addressing social inequality, discrimination, and aiding companies in attracting suitable candidates. Conventional surveys struggle to capture the nuances of online job searches that, as many online events, are characterized by repetition, low distinctiveness, and limited emotional impact. These characteristics can lead to memory-related errors, becoming more likely as the time between the event and the survey increases. Passively collected data, such as metered data provided by online panel members who install tracking software on their browsing devices, offer an alternative. While these data provide objective insights into online job searches, they suffer other types of errors, and cannot capture subjective information and all potential objective data of interest. This paper explores an alternative approach: sending surveys to individuals in a metered panel shortly after an event of interest is detected through metered data. These "in-the-moment" surveys aim to fill in missing information not obtainable through passive data collection while reducing memory-related errors that affect conventional surveys. To assess the feasibility and benefits of this method, an experiment comparing in-the-moment surveys triggered by online job applications with conventional surveys was conducted in an opt-in online panel in Spain to research how people apply for a job online. The results reveal that metered panelists accept well in-the-moment surveys, displaying high participation levels and positive evaluations regarding effort and satisfaction, without perceiving an increased privacy risk. Moreover, the data indicate positive impacts on data quality, with longer and more detailed responses to open-ended questions. However, not all aspects saw substantial improvements, with the reduction of non-recall being weaker than expected, possibly due to participants' overconfidence in their memories. The significant disparities observed in substantive results between both types of surveys also suggest that participants are not fully aware of what they do not remember.

Keywords: in-the-moment surveys, metered data, passive data, web surveys, digital traces, job applications.

The job market consistently ranks among citizens' top policy priorities in most countries,[1] as employment is vital for economic growth (Boltho & Glyn, 1995) and a fundamental aspect of mental health at the individual level (Ezzy, 1993). Within this field, job search is receiving increasing attention from researchers. Over the past decades, the internet has transformed job searching. By 2015, 54% of U.S. adults had researched jobs online (twice as many as in 2005), and 45% had applied for jobs online (Smith, 2015). Since then, the internet has become a crucial employment resource globally, but to an unequal extent for different population groups, such as older and less educated individuals. Therefore, understanding online job search and application behaviors is crucial, particularly for defining policies against social inequality and discrimination (Karaoglu & Hargittai, 2022) and for helping companies attract suitable candidates (Mansouri et al., 2018).

Several aspects of online job search have been investigated: effectiveness to escape unemployment (Kuhn & Mansour, 2014), impact of online reviews on job seekers (Faiz, 2020), use of different online platforms and their outcomes (Dillahunt, 2021), digital inequalities (Karaoglu & Hargittai, 2022) and gender differences (Fluchtmann et al., 2021).

However, research on job search is limited by the lack and/or inadequacy of available data. Most studies rely on surveys where participants report past job searches, which can be significantly affected by memory limitations. Job search involves a series of repetitive events (i.e., finding, reading, and applying for job offers) that are low in distinctiveness and emotional impact, involve little rehearsal (i.e., minimal time spent thinking/talking about each event), and are of short duration. These factors, combined with the passage of time, increase the likelihood of memory errors (Tourangeau, 2000) and recall bias (Walker & Skowronski, 2009). These issues are also prevalent in other online activities studied by researchers, such as housing searches, online purchases, and media consumption.

1 https://www.undp.org/sustainable-development-goals, https://www.ipsos.com/en/global-public-ranks-ending-hunger-and-poverty-and-ensuring-healthy-lives-top-priorities-among-un

Acknowledgements

I am very grateful to Melanie Revilla for her valuable and constant feedback and guidance in the development of this paper, and to her, Eva Fortes, Giampiero Passaretta, María José Gonzalez and Clara Cortina for their collaboration on the design of this research. Also, I would like to thank Patricia Iglesias for their review and comments on the questionnaire. This project received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No. 849165). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Direct correspondence to

Carlos Ochoa, Research and Expertise Centre for Survey Methodology,
Universitat Pompeu Fabra, Barcelona, Spain
E-mail: carlos.ochoa@upf.edu

Passively collected data, which do not require active participation in the data gathering from the observed individuals (Link et al., 2014), have also been used to study job seeking. For example, the professional social network Linkedin.com regularly releases job market reports based on data from job seekers and employers using their services. Similarly, opt-in online panels like Netquest and Yougov have requested some of their members who regularly participate in surveys to install tracking software on their browsing devices (a “meter”) to gather information on online activities, such as visited URLs, search terms, and app usage. Researchers can use these “metered” panels (Revilla et al., 2021) to investigate online job search behaviors.

In contrast to survey data, metered data (a type of digital trace data) are not subject to memory errors (Revilla, 2022), making them well-suited for collecting objective information, such as the number of job offers accessed per day or the time spent per offer. However, metered data can be affected by other errors (see “Background” section) and cannot capture subjective information, like why someone decides to apply for a job. Furthermore, meters cannot capture all objective information, such as whether the candidate secured the job.

This paper explores a third option: sending a survey to a sample of individuals from a metered panel when an event of interest is detected using metered data. This method can add missing information that cannot be collected passively, while reducing memory errors that affect retrospective surveys by shortening the time gap between the event and the data collection. However, some doubts arise about its applicability and effectiveness: (1) Will panelists agree to participate in such “in-the-moment surveys”? (2) How will respondents evaluate their experience? and (3) To what extent can these surveys provide better or new data compared to retrospective surveys?

This paper addresses these questions by reporting the results of an experiment comparing an in-the-moment survey triggered by online job applications detected through metered data with a conventional survey, i.e., a retrospective web survey sent to members of an opt-in online panel asking whether they applied for a job in the last six months. Both surveys requested additional information about one job application, along with sociodemographic and personality trait questions.

Background

Metered Data

Metered data offer substantial advantages over surveys for measuring online behaviors, such as greater granularity and robustness against memory errors, but are not error-free.

First, metered panels are usually formed from a subset of opt-in online panel members, which may introduce self-selection bias due to their non-probability-based recruitment (Baker et al., 2010). This bias can be exacerbated when panelists are asked to install a meter. Revilla et al. (2021), examining Netquest panels in nine countries, found that panelists who installed the meter when offered differed from those who did not in terms of gender, education, income, age, and panel loyalty. This self-selection may limit the capacity of metered panels to produce precise population estimates. Nonetheless, despite these limitations, opt-in online panels remain prominent in online research,[2] with metered data becoming increasingly popular in media, political, and social research (Revilla, 2022).

Besides representativeness issues, metered data suffer from other errors often overlooked by researchers, as discussed in Bosch and Revilla's (2022) Total Error framework for digital traces collected with Meters (TEM). When using metered data to trigger in-the-moment surveys, researchers must consider that these errors may cause the non-detection of events that should trigger surveys ("false negatives") and the detection of events that should not ("false positives;" Bosch et al., 2025).

False negatives and false positives can arise from various scenarios. False negatives include pausing the meter during job applications, technological limitations, inability to track mobile apps events, and overlooking relevant URLs. False positives can result from shared metered devices (Revilla et al., 2017), leading to incorrect event attribution. A key issue that can cause both false negatives and false positives, depending on the researcher's decisions, is when websites use the same URL for multiple events.

In summary, using metered data to detect events is fallible, potentially resulting in a sample that is not representative of all job applications. Additionally, non-detection of events extends the fieldwork time needed to reach the target sample size. Conversely, false positives require longer questionnaires with screening questions to exclude mistakenly detected participants, increasing data collection costs.

In-the-Moment Surveys Triggered by Metered Data

Participation

Previous research shows that metered panelists exhibit an overall high willingness to participate in in-the-moment surveys triggered by metered data, ranging from 69% to 95%, depending on the conditions offered to participants (Ochoa & Revilla, 2022). However, stated willingness may not always translate to actual

2 https://shop.esomar.org/knowledge-center/library-2021/Global-Market-Research-2020-pub2942

participation due to practical issues, such as not receiving or seeing the survey invitation in time (Ochoa & Revilla, 2022).

There has been little experimental research on in-the-moment surveys triggered by metered data, with one notable exception by Revilla and Ochoa (2018). In this study, a pop-up invitation was used to invite metered panelists of the Netquest panel in Spain to take part in a survey when a flight purchase was detected, but only 18 individuals completed it. The authors cited technological issues and short fieldwork times as possible reasons for the low participation. Overall, the limited evidence suggests that obtaining participants in the moment is a significant challenge.

Reduction of Recall Errors

The use of in-the-moment surveys aims to minimize the time gap between the event of interest and data collection, thus reducing errors from memory limitations. Tourangeau (2000) identifies four classes of memory problems: encoding issues (experiences not properly recorded), storage problems (corrupted memories), retrieval failures (inaccessible memories), and reconstruction errors (partially retrieved memories are inaccurately reconstructed).

When asking participants for subjective evaluations instead of factual data, such as reflections on feelings during an event, memory issues can worsen. These evaluations might not have been formed at the time, leaving nothing to remember, an extreme form of encoding problems. When asked later, reconstruction errors can occur if evaluations are made a posteriori, combining factual memories and present circumstances. This can be related to the discrepancy between the experiencing and remembering self (Kahneman & Riis, 2005).

The longer the time between an event and its recall, the greater the chance of retrieval and reconstruction failure. This applies to all types of events, from hospital stays to consumer purchases (Jobe et al., 1993). Most theories attribute the decline in accessibility over time to the interfering effects of later experiences, making online events (frequent and repetitive) particularly susceptible to rapid forgetting.

Some research methods leverage this fact to improve data quality. Ecological Momentary Assessment (EMA), for instance, prompts participants to report their current experiences via alarms sent at a predetermined or random schedule (Shiffman et al., 2008; van Berkel et al., 2017). This method avoids retrospective reporting but is impractical for studying specific events, as it would require frequent surveys to capture individuals experiencing the event of interest by chance. Coincidental surveys (Lamas, 2005) tried this approach in the early 20th century for measuring radio audiences but were found to be costly and operationally difficult.

In-the-moment surveys are similar to EMA, but target only those detected experiencing an event of interest. Besides metered data, other passive data sources can be used for detecting events and triggering surveys: GPS data (known as geofencing, e.g., Haas et al., 2020), smartphone accelerometer data (e.g., Hardeman, 2019), and Bluetooth beacons (e.g., Allurwar, 2016). All these methods combine self-reports with passively collected data (Keusch & Conrad, 2022).

However, even with these methods, a time gap between the event and the response may remain. While smaller than in retrospective surveys, this gap can still affect data quality.

Several empirical models have been proposed to quantify information loss over time (Rubin & Wenzel, 1996). They all predict that forgetting occurs rapidly at first and then slows down, supporting the benefits of in-the-moment surveys. However, since these “retention functions” may vary by individual and context, it remains unclear how close to the event surveys should be conducted to achieve a positive effect.

In this regard, Revilla and Ochoa (2018) compared survey responses collected up to 48 hours after the event of interest (probably too long to be considered “in the moment”) with up to two months later, finding no significant differences in answers.

Research Questions, Hypotheses, and Contribution

Based on the literature on in-the-moment research and its limitations, the following research questions and hypotheses are proposed.

RQ1. What are the levels of participation for in-the-moment surveys triggered by metered data among metered panelists compared to an equivalent conventional web survey?

Participation is expected to be slightly lower for in-the-moment surveys (*H1*). The technology used in this study, developed to detect online events and invite participants shortly after (see section “Software” in “Data and Methods”), aims to bridge the gap between the high willingness to participate reported in the literature and the low participation observed in the single previous study. However, in-the-moment surveys may interrupt participants, raise privacy concerns by highlighting the implications of sharing metered data, and be perceived as intrusive (Ochoa & Revilla, 2022), potentially decreasing participation.

RQ2. How do participants evaluate in-the-moment surveys compared to conventional web surveys?

In-the-moment surveys might be easier for participants, as they ask about fresh experiences and may seem more relevant. However, including questions unrelated to the event, such as sociodemographic ones, could dilute this positive effect. Additionally, the same issues affecting participation (see RQ1) may also impact evaluations. Thus, overall evaluations are expected to be similar to those of conventional surveys (*H2*).

RQ3. Is data quality higher from in-the-moment surveys compared to conventional web ones?

I expect in-the-moment surveys to produce better data quality (*H3*) due to reduced memory errors, such as fewer "don't remember/don't know" [3] responses (i.e., explicit non-recall). Additionally, lower effort required to answer event-related questions and higher respondent interest may reduce satisficing. This could lead to higher data quality, with longer and more meaningful responses to open-ended questions and fewer invalid and inconsistent answers.

Finally, some information (e.g., exact submission time of the job application) can be obtained from metered data for in-the-moment surveys, while conventional surveys require direct questions. This may lead to differences in non-response rates and answer precision.

RQ4. Do in-the-moment surveys lead to different substantive results compared to conventional surveys?

Both methods may yield different responses to the same questions due to varying sources of error and selection bias. Consequently, I expect different results for comparable questions related to the event of interest (*H4*). Assuming no other differences between the two methods except the time elapsed since the event (i.e., controlling for selection bias), responses given in the moment should have higher credibility. Therefore, discrepancies in substantive answers may indicate distorted recall (i.e., memory alterations of which respondents are unaware).

By addressing these research questions, this paper contributes to the existing knowledge in several ways. First, it explores the feasibility and potential benefits of in-the-moment surveys triggered by metered data, a topic not yet researched. Second, it tests a new approach to overcoming the technical issues that hindered the only previous academic attempt to develop such surveys. Finally, it evaluates in-the-moment surveys for studying real-world issues like job searching, which may inspire practical applications and provide insights into participants' perceptions of risks (e.g., privacy) and benefits in a real-world setting.

[3] While these answers do not convey the exact same meaning, in practice, they are often indistinguishable. Participants unable to provide an answer may not be aware if they did not see the requested information or if they forgot it. Both types of answers are reported separately in SOM4.

Data and Methods

Data

The data were collected from the Netquest opt-in online panel (www.netquest.com) in Spain. Netquest panel members regularly participate in surveys and earn points proportional to the length of the surveys, which can be redeemed for gifts (Revilla, 2017). In addition, some panelists are offered the possibility to install the meter in exchange for two to 12 additional points per week, depending on the number of devices where the meter is installed. The panelists invited to join the metered panel are not randomly selected from the survey panel. Instead, Netquest selects them based on their likelihood to accept the meter installation, determined by an internal predictive algorithm, and the need for participants for different research projects. The average installation rate is between 20% and 42%, depending on the country (Revilla et al., 2021).

Data for the in-the-moment surveys were collected from the 10th of March to the 3rd of October 2023 (207 days) using the metered panel. Data for the conventional survey were collected from the 30th of May to the 4th of June 2023 (five days) using the opt-in online panel, which includes the members of the metered panel.

In this study, the objective was to compare two samples of around 200 panelists who had applied for a job. A detailed description of how both samples were produced can be found in the "Methods" section, as this process is a fundamental part of this research.

Participants in the in-the-moment survey who did not confirm having applied for a job (105) and those who responded 48 hours after the application (21) were discarded. This decision was based on the results of Revilla and Ochoa (2018), who did not find relevant differences between responses collected 48 hours after the event and those collected up to two months later. Consequently, the final number of valid participants in this survey was 177, all of them metered panelists. Among them, 46.9% responded through a mobile device (smartphone or tablet). Their average age is 41.7 years, with 55.4% being women. 49.7% are mid-educated and 44.6% are highly educated. Their median number of participations in surveys in the last three months is 32.

As for the conventional survey, the number of valid participants was 201, out of which 56 were metered panelists, and the remaining 145 were participants in regular surveys only. 71.6% responded through a mobile device. The average age in this group is 38.6 years, with 61.2% being women. 47.3% are mid-educated and 44.8% are highly educated. Their median number of participations in surveys in the last three months is 25.

Both samples present significant differences in age, number of participations in surveys in the last three months, being metered and the device used to participate (see Appendix 1).

Software

Given past technical issues reported in the literature that made the implementation of in-the-moment surveys difficult, this study used WebdataNow (Revilla et al., 2022), software specifically designed for conducting in-the-moment surveys triggered by metered (or geolocation) data.

WebdataNow performs three main functions: (1) receiving metered (or geolocation) data from a panel, (2) identifying events of interest in the data, and (3) triggering survey invitations to the relevant panelists. The events of interest are defined by a list of regular expressions[4] that match the URLs intended to trigger the survey. Additionally, WebdataNow allows researchers to set a notification delay (the time between event detection and survey invitation) and a maximum time limit for participants to access the survey after the invitation is sent. For this study, the notification delay was set to five minutes.

Methods

To address the research questions raised in this study, the same topic (how individuals decide to apply for a job) was investigated using an in-the-moment and a conventional survey. These methods differ essentially in how candidate participants are selected and invited to participate. In addition, the questionnaire had to be adapted to each method. The following sections cover such differences.

Sample Selection

When selecting candidate panelists for the two samples, both metered and non-metered panelists were eligible for the conventional survey, while only metered panelists were eligible for the in-the-moment survey.

In this research, the opt-in online panel had fewer metered than non-metered panelists, risking the failure to reach the target sample size for the in-the-moment group. Prioritizing metered panelists for the in-the-moment group would have meant that the conventional survey sample consisted only of non-metered panelists. This would have led to two issues: first, it would not have provided a realistic sample for the conventional group, as the Netquest panel typically includes metered panelists in regular survey samples; second, it would have created a perfect correlation between the method (in-the-moment versus conventional) and the type of panelists (metered versus non-metered), hindering the identification of method-specific effects, which is the primary focus of this study.

[4] A regular expression is a sequence of characters that specifies a search pattern in text. See Appendix 2 for examples.

Therefore, I proceeded as follows. First, the in-the-moment survey was activated for a randomly selected half of the metered panelists, meaning that panelists who applied for jobs on one of the pre-identified websites (see subsection “In-the-moment survey”) using a device with the meter installed would receive an invitation to participate.

Second, once more than half of the in-the-moment target sample was achieved (30th of June 2023), non-metered and metered panelists who were not activated for the in-the-moment survey were randomly selected to form the sample for the conventional survey. The number of invitations sent was determined based on the target sample size (200) and the estimated proportion of panelists seeking a job (15%). This estimation was based on the number of visits to job search websites observed in the whole Netquest metered panel for Spain over a six months period.

Finally, after reaching the sample target for the conventional survey, the remaining non-invited metered panelists were activated for the in-the-moment survey. The detection of job applications stopped when the target sample size was reached (3rd of October 2023).

Following this process ensured that the conventional group included some metered panelists and that panelists were invited to participate in only one of the surveys.

In-the-Moment Survey

Panelists in the in-the-moment group received an invitation to participate in a survey five minutes after applying for a job on one of the listed websites (see Appendix 2). This invitation was sent only once during the project, the first time they were detected. This list covers the most popular job search platforms in Spain for which it was possible to identify a unique URL shown when a visitor applies for a job. Since the meter used in this research did not allow for the detection of activity occurring within apps, applications from apps could not be detected either.

The inability to detect all the participants’ job applications, together with other sources of error affecting metered data mentioned in the “Background” section, led to high levels of false negatives and false positives. Although not directly measurable, such levels are estimated to be around 85% and 34%, respectively (see SOM1).

All the detected panelists received the invitation by email. Additionally, panelists using the panel app also saw a push notification on their smartphones and/or tablets. The panel app, which facilitates various aspects of panel membership (e.g., invitation, redemption of points for incentives), can be installed voluntarily by the Netquest panelists, but is mandatory for those who want to install the meter on a mobile device (Revilla et al., 2021). As a result, approximately 90% of the metered panelists have the app.

Since the invitation to participate was sent via both emails and push notifications, participants did not necessarily take the survey on the same device where the job application was detected.

The message included in the invitation emphasized that participation time was limited, but without specifying a clear limit. However, the potential impact of this mention on participation was expected to be low since it could only be seen after opening the email and/or clicking on the app notification. We introduced a time limit message to encourage participants to provide responses promptly. Nevertheless, we allowed participants to complete the survey after this time limit to explore whether individuals would still participate, enabling us to potentially compare them with respondents who answered shortly after receiving the invitation. However, due to the limited sample size, a conclusive analysis comparing these two groups proved unfeasible.

Twenty-five percent of respondents completed the survey within 15 minutes after the job application and 50% within 72 minutes. However, the distribution of the delay in participating is strongly skewed to the right: 25% of respondents took more than 8 hours to participate, and 10% more than 2.3 days. For the analyses, we excluded responses submitted more than 48 hours after the invitation, as explained in the “Data” section.

Due to the possibility of shared devices, the questionnaire was designed to confirm that the panelist was indeed the one who applied for the job, without disclosing private information in case it had been obtained from a third party. To achieve this, participants were asked, after obtaining informed consent, whether they had engaged in four different online activities within the last 48 hours. One of these was “reading job offers”. Only those who responded affirmatively to this question were allowed to proceed with the questionnaire. By adding this step, the risk of revealing third-party private information to the participant and causing any harm was considered to be extremely low.

Then, the questionnaire explicitly informed participants that they were invited to participate because they were detected looking at a job description, with the specific website and approximate time of detection provided. Approximately half of the sample was informed that the survey was sent close to the event of interest to enhance the quality of the data for researchers, while the other half was told that the purpose was to help them recalling their answers more easily. This message aimed to assess if the communicated benefit of the method had an impact on the results.[5]

Participants were then asked to confirm whether they had visited the job offer and whether they had finally applied for the job. After this section, the questions used for both the substantive and methodological research were presented. An

---

5 The results of this experiment indicated a slight and non-significant effect in favor of communicating that the main benefit is for the respondent in terms of breakoff and survey evaluation. For more details, see SOM4.

English translation of the full questionnaire and screenshots are available in the supplementary online material (SOM2).

The questionnaire aimed to assess potential differences in online job application behavior among various demographic groups (especially males and females), including whether participants met all job requirements, their self-reported likelihood of being hired, and whether the job position met their expectations. Additionally, the questionnaire included questions about participants' sociodemographic background and personality traits (19 items in two batteries of questions). One of the key substantive hypotheses that this survey aimed to confirm was whether females applied for a job offer less frequently than males when they did not meet all the requirements of the job offer (Ochoa et al., 2023).

Five questions to assess participant's evaluation of the survey were also asked in the questionnaire, before the personality trait questions.

The full questionnaire included up to 69 questions and was optimized for mobile devices. The average time to complete it was 10.2 minutes and the median 9.2 minutes. Respondents could continue without answering the questions, except those used to filter other questions. A warning message was shown to 7.9% of participants who tried to skip a question when multiple questions were presented on the same page. Following the panel's usual practice, going back was not allowed.

Conventional Survey

The in-the-moment questionnaire was adapted to be used in a conventional survey. Since in a conventional survey it is not possible to refer to a concrete job application detected using metered data, participants were asked about their most recent job application in the last six months. Questions such as "Why did you apply for this job offer?" were rephrased as "Think about the last job application you submitted online for a job offer. Why did you apply to this job offer?"

Besides reformulating job application related questions, other changes were made:

- The initial section designed to verify that the panelist was the one who made the job application was removed.
- Two questions were added to gather when and in which website the application took place. In the in-the-moment survey this information was gathered using metered data.
- A question was added asking participants to what extent they were confident (0 to 100%) that the job application they reported was actually the last one they did.

The final conventional questionnaire (see SOM2) included up to 69 questions. The average time to complete the questionnaire was 9.6 minutes and the median 8.6 minutes. All the remaining features of the questionnaire (e.g., possibility to skip questions, etc.) were the same as in the in-the-moment questionnaire. The warning message presented when trying to skip a question in pages with multiple questions was shown to 11.4% of participants.

Analyses

Comparisons Between Groups

The analyses were performed using R version 4.2.3. Various metrics (participation metrics, survey evaluations, quality indicators, and substantive answers) are calculated for participants in the in-the-moment group and compared with those in the conventional group.

When the calculated metrics represent proportions (e.g., proportion of participants evaluating the survey as "easy"), Fisher exact tests are used for group comparisons. For metrics representing means (e.g., mean number of characters in answers to an open-ended question), *t*-tests are used for comparisons. Metrics representing means of reported percentages or probabilities (e.g., estimated probability of being hired) are compared also using *t*-tests. In all these cases the resulting *p*-values are reported.

As described in the "Data" section, the sample selection method did not guarantee equal sample compositions in both groups. To account for these differences, logistic regressions are conducted for dichotomous variables, and linear regressions for continuous numerical variables, while controlling for three sociodemographic variables: gender (two groups), age (numeric), and education level (two groups). Additionally, two panel variables are used as controls: the number of participations in Netquest surveys in the three months before this study[6], and being a metered panelist. The inclusion of this last variable is crucial because all participants in the in-the-moment group are metered, whereas only 28% of the conventional group participants are. This factor could potentially confound both the method and sample composition effects in a direct comparison. Similarly, the type of device used to complete the survey, which may influence data quality (Lambert & Miller, 2015), was included as a control variable because the proportion of PCs is significantly larger in the in-the-moment group (see Appendix 1).

Similar to the direct comparisons, *p*-values are reported for the regression analyses, using a significance level of 5% in both cases. However, due to the limited sample size, detecting significant effects with all these covariates poses challenges, especially for questions presented to only a subset of respondents due to the questionnaire's routing conditions.

---

6 I also attempted including the total number of participations in panel surveys and the log transformation of both variables as covariates, which yielded comparable results. Details of these analyses can be found in the SOM4.

Open Questions

Several open-ended questions were used to gather both objective (e.g., name of the employer) and subjective (e.g., main reason to switch jobs) information about the job applications. Answers to such questions required coding to capture their substantive meaning while also assessing data quality, including non-recall, off-topic answers, or overall answer coherence.

Answers were coded by two native speakers. Initially, the main coder created a codebook. Then, a secondary coder used the same codebook to repeat the process. The intercoder reliability was 96%. The reported results are those produced by the main coder, after reviewing those of the secondary coder.

Participation

To compare participation levels between the two groups, three primary metrics are used: (1) Start rate, which indicates the proportion of invited panelists who initiate the survey (starts) compared to the total number of invited panelists (invites). (2) Breakoff rate, which represents the percentage of panelists who abandon the survey (breakoffs) divided by the number of panelists who start the survey (starts). (3) Incidence rate, calculated as the number of valid surveys (completed surveys not discarded due to screening questions) over the total number of completed surveys (completes).

The start rate helps assessing whether inviting panelists while they are actively engaged in the activity of interest (job search) leads to a higher likelihood of them disregarding the survey invitation. Conversely, the dropout rate provides insights into whether inquiring about a recent meter-detected activity prompts participants to abandon the survey less frequently without completing it. Lastly, the incidence rate, which measures fieldwork efficiency (Ochoa & Porcar, 2018), assesses the potential benefits of contacting people in the moment in terms of sample utilization.

Survey Evaluations

Five questions are used to evaluate participants' perceptions of both surveys: self-reported effort to participate, satisfaction, trust in survey anonymity, perceived intrusiveness, and willingness to participate again in a similar survey.

The first four questions utilized scales with five levels, consisting of two negative options (e.g., very difficult, quite difficult), one neutral option (e.g., neither difficult nor easy), and two positive options (e.g., quite easy, very easy). For each of these questions, the proportion of positive answers, combining the two positive levels, is compared.

The question regarding willingness to participate again involved three response options (yes, no, and not sure). The proportion of affirmative answers between the two surveys is compared.

Data Quality

To assess differences in data quality between the groups, five commonly used indicators are employed. The full details of the variables used for each indicator can be found in Appendix 3. The indicators are:

1. *(Explicit) non-recall*: This indicator measures the proportion of respondents unable to recall requested information in a question, attributed to the effect of time and/or the lack of effort (Groves, 1989). I focus on explicit non-recall, where participants overtly declare their inability to provide the requested information. This evaluation spans across 22 different questions, including open-ended questions and questions with “Don’t know” and/or “Don’t remember” options (both considered as non-recall). Two of the open-ended questions for the conventional group were not asked to the in-the-moment group, as the same information was obtained through metered data.
2. *Invalid answer*: Invalid answers, which serve as an indicator of low data quality (Revilla & Ochoa, 2015), were identified through manual coding of responses to eight open-ended questions. A response is considered invalid if it fails to answer what was asked.
3. *Length of answers*: The mean number of characters in the answers to narrative open-ended questions, after discarding invalid answers, is sometimes used as a measure of data quality (Revilla & Ochoa, 2015). This indicator is calculated for three different open-ended questions. Two additional questions were excluded from the analysis due to a very limited number of responses (less than 20 per group).
4. *Straight-lining*: Straight-lining refers to selecting the same option in a set of consecutive questions sharing the same answer scale, even when it is not reasonable to expect identical responses (Green & Krosnick, 2001). This indicator is calculated for one set of four questions and another set of eight questions.
5. *Inconsistencies*: Inconsistencies are assessed by analyzing the proportion of answers to specific questions where participants do not adhere to the instructions or provide combinations of answers that do not logically align (DiLalla & Dollinger, 2006), considering three groups of cases:
   - Numerical answers out of bounds for four open-ended numerical questions. Inconsistencies are noted when participants provide answers outside the range of 0-100%.
   - Incoherent answers across three groups of related questions, where the answer to one question should logically align with the answer to another question (e.g., the number of applications without meeting requirements should be below the total number of applications).

– Selecting more than the maximum allowed in a multiple-answer question.

Certain potential indicators, such as survey duration, were discarded due to their unclear relationship with quality, especially in online surveys where respondents may keep the survey open but inactive while engaging in other activities. Moreover, technical limitations hindered the utilization of some indicators, such as assessing the external validity of answers by comparing them to the actual job description seen by participants. Future versions of the meter may address this limitation.

Differences in Substantive Results

The potential effect of the survey type on substantive answers is assessed by comparing results derived from six questions requesting objective information (e.g., the percentage of met requirements) and two requesting subjective information (e.g., the expected probability of being interviewed).

Additionally, as control measures, two substantive results completely unrelated to the event of interest (personality traits that should not present differences between groups) are explored. The full detail on which variables are used for the substantive results can also be found in Appendix 3.

Results

Participation (RQ1)

Table 1 provides a summary of participation levels in both surveys. The percentages in the table are calculated relative to the preceding category, as indicated by the indentation of these categories in the first column. For instance, the percentage of starts for the in-the-moment survey (88.2%) is derived from the number of invited panelists. Similarly, the percentage of breakoffs (1.3%) is based on the number of starts, and so forth.

The ratio of participants who initiated the survey over the total number of invited panelists is significantly higher for the in-the-moment group (88.2%) compared to the conventional group (62.5%). When accounting for the 283 panelists from the conventional group who attempted to start the survey but found it closed due to reaching the target sample size ("Survey closed" in the table, 13.6%), the overall figure increases to 76.1%, which is still significantly lower than that of the in-the-moment group. Similarly, the percentage of breakoffs is significantly lower in the in-the-moment group (1.3%) compared to the conventional one (5.0%).

Table 1 Participation in in-the-moment (ITM) and conventional (Conv) surveys

	ITM		Conv	
	n	%	*n*	%
Invited	356		2,080	
Non-starts	42	11.8	498	23.9
Starts	314	88.2	1,299	62.5
Breakoffs	4	1.3	65	5.0
Non-consent	5	1.6	58	4.5
Screened-out	107	34.1	975	75.1
Not searching in the last 48h / 6 months	30	28.0	791	81.1
Not confirming last search / –	24	22.4	-	-
Not applying to the detected job / any job	51	47.7	173	17.7
Other (e.g., bot detection)	2	1.9	11	1.1
Complete	177	56.4	201	15.5
Complete after 48h	21	6.7	-	-
Survey closed	-	-	283	13.6

It is worth noting that these differences can be attributed to both the survey type and the profile of participants in each group. Specifically, while all participants invited to the in-the-moment survey are metered panelists, only 467 out of the 2,080 who started the conventional survey (22.5%) fall into this category. Metered panelists, who may generally have a more positive attitude toward surveys (Revilla et al., 2021), could contribute to the higher level of participation and lower level of breakoff rates observed in the in-the-moment group.

To further explore these effects and disentangle the impact of the type of survey from the panelists' profile, logistic regression analyses are conducted with participation and breakoff as the dependent variables, and survey type as the main independent variable. The analyses also controlled for the variables detailed in the section "Comparisons Between Groups" (see Appendix 1 for a descriptive analyses per group).[7]

Given the strong correlation between survey type and metering status ($r = .75$), a multicollinearity analysis was conducted. The Variance Inflation Factor (VIF) values for these two variables indicate acceptable levels of multicollinearity (see SOM3). Therefore, the specification of the model using all seven aforementioned variables was retained.

7 The type of device used to complete the survey was excluded from the participation analysis since this variable is only recorded once participants start the survey.

After controlling for these variables, the positive effect of the in-the-moment survey on participation remains significant ($p < .001$), but not on breakoff, where gender, age, and, specially, the number of past participations account for more explanatory power. However, these results lead us to reject the hypothesis that in-the-moment surveys have lower participation levels than conventional surveys (*H1*).

It is noteworthy also to discuss the variations in sample utilization between the two methods. To obtain 177 complete surveys for the in-the-moment group, a total of 315 panelists were invited, and among those who participated, 57.1% yielded valid participations (incidence rate). Conversely, for the conventional group, a significantly larger number of 2,080 panelists were invited, and only 16.3% of the participants ultimately provided valid responses.[8]

The difference in sample utilization can primarily be attributed to the need to ask participants in the conventional survey whether they applied for a job in the last six months, something that is known in advance for most of the in-the-moment survey participants. However, the improved sample utilization in the in-the-moment group comes at the cost of an extended fieldwork period (207 days vs. five days).

Survey Evaluations (RQ2)

Table 2 presents a comparison of survey evaluations made by participants in each group. Sample sizes are provided in Columns 2 and 3. Columns 4 to 7 display the proportions of positive answers for the in-the-moment group (ITM) and the conventional group (Conv), along with the difference (Diff = ITM-Conv) and the *p*-value resulting from a significance test. The last two columns present the impact of the in-the-moment group in a logistic regression model, incorporating the control variables described before.

Participants in the in-the-moment survey perceive the assigned task as significantly easier (+10.2 percentage points, pp) and more satisfactory (+11.6 pp) compared to the conventional survey. However, levels of trust in survey confidentiality and perceived intrusiveness are similar between the two surveys. This suggests that panelists do not perceive any additional risk in participating in the in-the-moment survey, despite the explicit mention of the invitation being triggered by activity detected using metered data. Moreover, the willingness to participate again is similarly high (94.4% and 93.5%).

[8] This incidence rate aligns with our initial estimations based on metered data, which further supports the notion that metered and non-metered panelists exhibit similar behavior, at least regarding online job applications.

Table 2 Survey evaluations by group

	N		Proportion %				Log. regression	
Question	ITM	Conv	ITM	Conv	Diff	*p*-value	Effect	*p*-value
Effort: easy	177	201	85.3	75.1	10.2*	.015	0.6	.121
Satisfaction: high	177	200	70.1	58.5	11.6*	.024	0.3	.345
Privacy: trust	177	200	74.4	70.0	4.4	.358	0.3	.474
Intrusiveness: low	177	201	50.3	52.2	-2.0	.757	-0.3	.431
Do it again: yes	177	200	94.4	93.5	0.9	.831	0.3	.582

Notes: Sample sizes for the in-the-moment (ITM) and conventional (Conv) groups. Proportion %: percentage of positive answers per group, difference in proportions (Diff) and significance (*p*-value). Log. regression: coefficient (Effect) and significance (*p*-value) of the in-the-moment group in a logistic regression controlling for gender, age, education level, metering status, number of participations (last three months) and device used to participate.

Once again, these differences appear to be a combined effect of the survey type and sample profile. The percentage of participants who rated their participation as easy was 74.5% for non-metered conventional participants, 76.8% for conventional metered panelists, and 85.3% for in-the-moment (metered) participants. Similarly, the percentages of participants who reported liking the participation experience in these three groups were 56.6%, 62.5%, and 70.1%, respectively. However, when conducting regression analyses controlling for being a metered panelist along with sociodemographic variables, the effect of the survey type is no longer statistically significant, which may be due to the limited statistical power resulting from splitting the sample into these groups.

In conclusion, despite the limitations posed by the smaller sample size, it can be inferred that in-the-moment surveys receive similar evaluations in terms of ease and satisfaction, compared to conventional surveys (support for *H2*).

Differences in Data Quality (RQ3)

The results of evaluating the 43 quality indicators described in Appendix 3 are presented in Table 3, following exactly the same structure as in Table 2.

Table 3 Quality indicators

	N		% of cases				Log. regression	
Non-recall indicators	ITM	Conv	ITM	Conv	Diff	*p*-value	Effect	*p*-value
Company name	177	201	16.4	25.9	-9.5*	.032	-0.6	.117
Job description	177	201	2.3	10.4	-8.2*	.001	-1.4*	.039
Salary	177	201	13.6	20.4	-6.8	.101	0.0	.952
Contract	177	201	17.5	25.4	-7.9	.080	0.1	.817
Experience	177	201	14.1	23.4	-9.3*	.026	-0.6	.173
Perks	177	201	14.7	20.9	-6.2	.140	-0.7	.058
% of met requirements	177	201	23.7	19.4	4.3	.318	0.2	.601
Specific not met req.	97	101	7.2	6.9	0.3	1	0.3	.735
% of fit	177	201	9.6	18.9	-9.3*	.013	-0.8	.075
Salary – Not fitting	134	133	7.5	3.8	3.7	.288	1.0	.346
Hours – Not fitting	134	133	1.5	0.8	0.7	1	18.1	.998
Flexibility – Not fitting	134	133	1.5	3	-1.5	.447	15.1	.993
Location – Not fitting	133	133	1.5	0.8	0.8	1	-0.6	.643
Tasks – Not fitting	134	133	0.0	0.0	0.0	1	0.0	1
Manager – Not fitting	134	133	2.2	2.3	0.0	1	15.6	.993
Company – Not fitting	134	133	2.2	3.8	-1.5	.500	15.4	.993
Contract – Not fitting	134	133	6.0	4.5	1.5	.785	16.6	.992
Applications in last 6m.	177	201	32.2	34.3	-2.1	.743	-0.2	.538
Apps. without req. in last 6m.	176	201	37.5	44.8	-7.3	.173	-0.5	.111
Apps. without fit in last 6m.	176	201	36.9	40.8	-3.9	.460	-0.6	.069
Job search website	177	201	-	2.5	-2.5*	-	-*	-
Last application date	177	201	-	49.8	-49.8*	-	-*	-
	N		% of cases				Log. regression	
Invalid answers	ITM	Conv	ITM	Conv	Diff	*p*-value	Effect	*p*-value
Company name	177	196	0.6	1.5	-1.0	.625	14.9	.995
Job description	177	195	1.1	3.6	-2.5	.178	-1.2	.234
Salary	51	54	0	13.0	-13.0*	.013	-19.1	.994
Contract	89	106	0	1.9	-1.9	.501	-0.3	1
Experience	106	103	0	2.9	-2.9	.118	0.8	1
Perks	17	12	17.6	16.7	1.0	1	95.4	1
Why applying without req.	77	91	5.3	11.0	-5.8	.263	-1.0	.220
Why applying without fit	134	132	16.7	21.4	-5.6	.350	-0.9*	.027

Table 3 (continued)

	N		Num. of characters.				Lin. regression	
Length of answers	ITM	Conv	ITM	Conv	Diff	*p*-value	Effect	*p*-value
Job description	175	188	41.3	28.0	13.3*	<.001	11.7*	.012
Why applying without req.	74	81	71.4	52.0	19.5*	.004	21.5*	.047
Why applying without fit	112	103	60.8	54.6	6.1	.325	11.2	.263
	N		% of cases				Log. regression	
Straight-lining	ITM	Conv	ITM	Conv	Diff	*p*-value	Effect	*p*-value
Job details (4 items)	177	201	10.7	15.4	-4.7	.224	-0.7	.103
Fit of features (8 items)	134	133	9.0	9.8	-0.8	.837	-0.4	.490
	N		% of cases				Log. regression	
Inconsistencies	ITM	Conv	ITM	Conv	Diff	*p*-value	Effect	*p*-value
% of req. out of limits	135	162	0	0	-	-	-	-
% of fit. out of limits	160	163	0	0	-	-	-	-
Probability of interview out of limits	176	200	0	0	-	-	-	-
Probability of hiring out of limits	176	200	0	0	-	-	-	-
% of met req. < 100 + meeting all req.	135	162	14.1	6.2	7.9*	.030	0.0	.965
> 3 options in motivation question	177	201	7.3	6.0	1.4	.680	-0.1	.848
Apps. without met req. > total apps.	104	105	1.9	1.9	0.0	1	-0.7	.571
Apps. without perfect fit > total apps.	103	112	3.9	1.8	2.1	.430	-0.8	.377

Notes: Sample sizes for the in-the-moment (ITM) and conventional (Conv) groups. Proportion %: percentage of positive answers per group, difference in proportions (Diff) and significance (*p*-value). Log./Lin. regression: coefficient (Effect) and significance (*p*-value) of the in-the-moment group in a regression (linear for means, logistic for proportions) controlling for gender, age, education level, metering status, number of participations (last three months) and device used to participate.

Out of the 22 non-recall indicators, 14 show better results for the in-the-moment group, indicating lower non-recall (negative effects). However, only six of these effects are significant. When adding the control variables, the number of favorable results for the in-the-moment groups decreases to 11, with three of them significant.

The largest favorable effect is observed for one of the two variables recorded using metered data instead of relying on a question for the in-the-moment group, with a decrease of 49.8 pp. Excluding this variable, the observed effects range from –9.5 pp to +4.3 pp. The median effect across all 22 variables is –2.3 pp, while the mean effect is –5.2 pp.

Among the six questions that exhibit higher levels of non-recall for the in-the-moment group, four belong to the same set of eight questions that asked participants whether each job feature matched what they were looking for, with an explicit option of "I don't remember." Interestingly, one of the other two questions showing a higher level of non-recall for the in-the-moment group is the one asking for the percentage of met requirements. This question included an input box for participants to write their answer and two radio buttons to indicate "I don't know" and "I don't remember" (see questionnaire in SOM2). In Table 3, both options are considered as non-recall. However, if only the option "I don't remember" is considered as non-recall, the in-the-moment group exhibits a lower level of non-recall (8.8% vs. 12.0%; see SOM4).

In terms of the percentage of invalid answers, seven out of eight indicators studied show lower levels for the in-the-moment group, but the effects are generally moderate, only one of them being significant. When controlling for the usual variables, the significant effect remains but two effects are reversed.

Regarding the length of answers to open narrative questions, all three questions studied show longer answers for the in-the-moment group, with relative effects ranging from +11.4% to +47.5%. Two of these effects are significant, also when controlling for the usual covariates.

The two straight-lining indicators favor the in-the-moment survey, although the effect is not statistically significant. However, this effect is only substantial (–4.7%) in the case of the set of four questions.

Finally, out of the eight consistency indicators, the four related to exceeding the limits of numerical questions do not show a single case in any of the two groups, while the remaining four exhibit very small and non-significant differences, with slightly worse results for the in-the-moment group.

In conclusion, these results do not clearly support the beneficial effects of in-the-moment surveying on data quality (*H3*), except for longer answers to open-ended questions. While statistically significant effects are lacking, 25 out of 44 indicators favor in-the-moment surveys, 12 favor conventional surveys, and 7 are neutral. This favorable trend for in-the-moment surveys needs further validation with larger samples. The positive impact on response length contrasts with the weaker-than-expected effect on non-recall, possibly due to participants' overconfidence in their memory accuracy.

To assess this potential overconfidence, participants in the conventional survey were asked to what extent they were confident that the information they reported actually corresponded to their last job application, as requested. Given that memories tend to fade over time, particularly in the initial stages, one might expect participants to report lower confidence levels for job applications made further in the past. However, as depicted in Figure 1, reported confidence levels remain relatively constant and consistently high across the 0 to 160-day range. This contradicts existing knowledge about memory decay, suggesting thus that overconfidence is occurring.

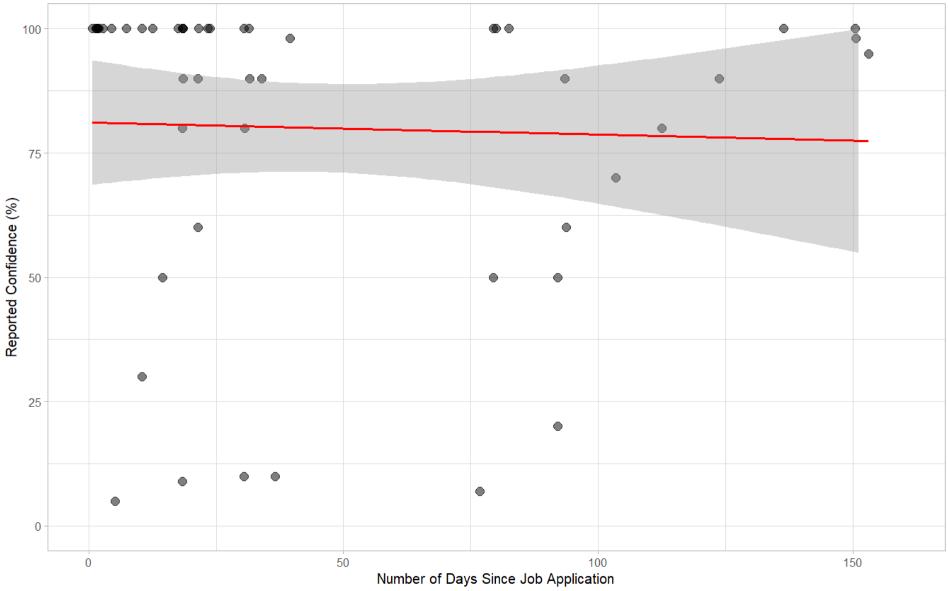

Figure 1 Confidence in reporting last job application over time by participants in the conventional survey

Differences in Substantive Results (RQ4)

Table 4 presents the substantive answers of interest for the two groups, maintaining the same structure as previous tables.

Despite the limited sample size, which particularly affects some questions asked only to part of the respondents, several significant disparities emerge in the substantive findings depending on the survey type. The estimated probabilities of being interviewed and hired, both subjective measures, are 7.9 pp and 8.7 pp lower, respectively, for the in-the-moment group, both significant. After adjusting for the usual covariates, the effects become 7.8 pp and 10.9 pp, with only thelatter remaining significant. As the questions in the conventional survey explicitly requested participants to report information "at the time of applying," the observed differences can be attributed to distorted recall, meaning recall errors or alterations introduced in the answers without participants being aware.

Table 4 Substantive differences

	N		Value (%)				Regression	
	ITM	Conv	ITM	Conv	Diff	*p*-value	Effect	*p*-value
Met requirements	135	162	80.1	84.0	-3.9	.082	-4.6	.162
Non-compliants	135	162	71.9	62.3	9.6	.108	0.4	.297
Fit of features	160	163	72.3	76.9	-4.6*	.046	-3.4	.340
Non-fitters	160	163	83.8	81.6	2.2	.660	0.3	.568
% apps. without met req	88	92	52.8	46.3	6.5	.242	17.5*	.046
% apps. without perfect fit	88	92	48.9	48.4	0.5	.933	-0.9	.915
Prob. of interview	176	200	47.7	55.6	-7.9*	.006	-7.8	.074
Prob. of hiring	176	200	39.6	48.3	-8.7*	.003	-10.9*	.012
Conformity (control)	177	201	2.7	2.7	0.0	.838	-0.1	.133
Efficacy (control)	176	201	3.9	3.9	0.0	.680	0.0	.836

Notes: Sample sizes for the in-the-moment (ITM) and conventional (Conv) groups. Value: value of the substantive result (a proportion or an average numeric value) per group, difference of values (Diff) and significance (*p*-value). Regression: coefficient (Effect) and significance (*p*-value) of the in-the-moment group in a regression (linear for means, logistic for proportions) controlling for gender, age, education level, metering status, number of participations (last three months) and device used to participate.

The percentage of participants who admitted to applying without meeting all the job requirements, a key aspect that motivated this research, especially in terms of potential gender differences, differs between the two survey types: 71.9% for the in-the-moment survey versus 62.3% for the conventional survey (on average 23.6 days after the event of interest). However, this effect vanishes when controlling for the usual covariates.

In contrast, the two personality traits included as controls yield almost identical results in both groups (with and without controls), aligning with our expectations. Therefore, these findings support the hypothesis (*H4*) that the time elapsed since the occurrence of the event of interest can impact the substantive answers provided by survey participants.

Discussion

Summary of Main Results

The results from this study reveal that members from metered panels readily embrace in-the-moment surveys triggered by online events, displaying heightened levels of participation in comparison to an equivalent conventional survey (RQ1). Furthermore, participants evaluated this new method similarly to a conventional survey regarding the required effort and overall satisfaction, and did not perceive an increased risk concerning privacy or intrusiveness associated with this survey format (RQ2).

The responses provided by participants also revealed moderate positive impacts on data quality (RQ3). While some of these effects are substantial and statistically significant, such as the increase in the length of responses to open-ended questions, other observed effects are not conclusive. Particularly, the positive effect on explicit non-recall is weaker than expected, possibly due to participants' overconfidence in the accuracy of their memories. Moreover, significant disparities in substantive results emerged based on the data collection method employed, further supporting the notion that recall errors influence the gathered data more extensively than participants are aware (RQ4).

Limitations

There are several limitations affecting this study. Firstly, it relies on a sample from a single opt-in online panel (Netquest) in a single country (Spain). Different panels and countries, as well as other sampling methods (e.g., probability samples) may produce different results, underscoring the need for caution when attempting to generalize findings.

Secondly, the technical solutions chosen for the study might have influenced the outcomes. Utilizing alternative platforms could lead to different results. For instance, the way push notifications are presented to mobile panelists can significantly impact their participation behavior (e.g., shorter/longer delays). Similarly, the impossibility of detecting certain online events (i.e., events within mobile apps) may have affected some results.

Thirdly, the sample size for this research was constrained by the availability of metered panelists, preventing certain analysis (e.g., the effect of elapsed time since the job application on discrepancies in substantive results between both surveys; see SOM4) potentially limiting the ability to detect significant effects for certain observed differences. To validate the findings of this paper, further investigations with larger sample sizes are necessary, maybe focusing on high-prevalence events (e.g., online purchases) to address the current limitations stemming from the constrained size of existing metered panels.

Practical Implications

Surveying people in the moment using metered panels is a promising methodology, especially suited for researching repetitive, low-emotional, and hard-to-distinguish events. Despite the limited sample size in this study, the results indicate that conducting research close to the event of interest leads to slightly better data quality and reveals clear differences in substantive results. Such substantive differences suggest a significant reduction of distorted recall, wherein people inadvertently fail to report accurate information.

Nevertheless, this study has also highlighted several inconveniences associated with this type of surveys that researchers must carefully consider before deciding to use it. In-the-moment surveys require the use of specific technology, a complex set-up that involves the identification of specific URLs related to the events of interest, extended fieldwork times, and regular supervision, among other challenges (see Reflective Appendix). Additionally, the limited size of existing metered panels poses limitations on obtaining large samples for specific target populations.

This combination of pros and cons suggests that in-the-moment surveys are well-suited for high prevalence activities that occur frequently over time, but they may not be ideal for activities with an excessive number of repetitions in a short period. The latter scenario could lead to participants misidentifying the specific event of interest in the survey. Examples of suitable activities may include post-purchase satisfaction surveys for online purchases and opinion polls targeting audiences during live streaming media consumption.

Finally, it is essential to emphasize that in-the-moment surveys are not designed to replace conventional surveys but rather to serve as a valuable additional methodology in very specific cases. Their unique strengths make them particularly useful for certain research scenarios, at the cost of extended fieldwork times and increased complexity.

Data Availability and Supplementary Online Material (SOM)

The anonymized dataset, together with all the scripts used for the analyses and the supplementary online material of this paper can be found at: https://osf.io/67sgz

References

Allurwar, N., Nawale, B., & Patel, S. (2016). Beacon for proximity target marketing. *International Journal of Engineering and Computer Science, 15*(5), 16359–16364. https://ijecs.in/index.php/ijecs/article/view/1125

Baker, R., Blumberg, S. J., Brick, J. M., Couper, M. P., Courtright, M., Dennis, J. M., Dillman, D., Frankel, M. R., Garland, P., Groves, R. M., Kennedy, C., Krosnick, J., Lavrakas, P. J., Lee, S., Link, M., Piekarski, L., Rao, K., Thomas, R. K., & Zahs, D. (2010). Research synthesis: AAPOR report on online panels. *Public Opinion Quarterly, 74*(4), 711–781. https://doi.org/10.1093/poq/nfq048

Boltho, A., & Glyn, A. (1995). Can macroeconomic policies raise employment? *International Labour Review, 134*(4–5), 451–470.

Bosch, O. J., & Revilla, M. (2022). When survey science met web tracking: Presenting an error framework for metered data. *Journal of the Royal Statistical Society Series A: Statistics in Society, 185*(Supplement_2), S408–S436. https://doi.org/10.1111/rssa.12956

Bosch, O. J., Sturgis, P., Kuha, J., & Revilla, M. (2025). Uncovering digital trace data biases: Tracking undercoverage in web tracking data. *Communication Methods and Measures, 19*(2), 157–177. https://doi.org/10.1080/19312458.2024.2393165

DiLalla, D. L., & Dollinger, S. J. (2006). Cleaning up data and running preliminary analyses. In F. T. L. Leong & J. T. Austin (Eds.), *The psychology research handbook: A guide for graduate students and research assistants* (2 ed., pp. 241–254). Sage Publications. http://doi.org/10.4135/9781412976626.n16

Dillahunt, T. R., Israni, A., Lu, A. J., Cai, M., & Hsiao, J. C. Y. (2021). Examining the use of online platforms for employment: A survey of US job seekers. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems,* Article 562, 1–23. https://doi.org/10.1145/3411764.3445350

Ezzy, D. (1993). Unemployment and mental health: A critical review. *Social Science & Medicine*, *37*(1), 41–52. https://doi.org/10.1016/0277-9536(93)90316-V

Faiz A. (2020). Impact of online job search and job reviews on job decision. *Proceedings of the 13th International Conference on Web Search and Data Mining,* 909–910. https://doi.org/10.1145/3336191.3372184

Fluchtmann, J., Glenny, A., Harmon, N. A., & Maibom, J. (2022). *The gender application gap: Do men and women apply for the same jobs?* (IZA Discussion Paper No. 14906). SSRN. https://ssrn.com/abstract=4114410

Green, M. C., & Krosnick, J. A. (2001). Comparing telephone and face-to-face interviewing in terms of data quality: The 1982 national election studies method comparison project. In M. L. Cynamon & R. A. Kulka (Eds.), *Health survey research methods* (pp. 115–121). U.S. Department of Health and Human Services.

Groves, R. M. (1989). *Survey errors and survey costs*. John Wiley & Sons.

Haas, G.-C., Trappmann, M., Keusch, F., Bähr, S., & Kreuter, F. (2020). Using geofences to collect survey data: Lessons learned from the IAB-SMART study [Special issue: Advancements in online and mobile survey methods]. *Survey Methods: Insights from the Field*. https://doi.org/10.13094/SMIF-2020-00023

Hardeman, W., Houghton, J., Lane, K., Jones, A., & Naughton, F. (2019). A systematic review of just-in-time adaptive interventions (JITAIs) to promote physical activity. *International Journal of Behavioral Nutrition and Physical Activity, 16,* Article 31. https://doi.org/10.1186/s12966-019-0792-7

Jobe, J. B., Tourangeau, R., & Smith, A.F. (1993). Contributions of survey research to the understanding of memory. *Applied Cognitive Psychology, 7*(7), 567–584. https://doi.org/10.1002/acp.2350070703

Kahneman, D., & Riis, J. (2005). Living, and thinking about it: Two perspectives on life. In F. A. Huppert, N. Baylis, & B. Keverne (Eds.), *The science of well-being* (pp. 285–304). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198567523.003.0011

Karaoglu, G., Hargittai, E., & Nguyen, M. H. (2022). Inequality in online job searching in the age of social media. *Information, Communication & Society, 25*(12), 1826–1844. https://doi.org/10.1080/1369118X.2021.1897150

Keusch, F., & Conrad, F. G. (2022). Using smartphones to capture and combine self-reports and passively measured behavior in social research. *Journal of Survey Statistics and Methodology, 10*(4), 863–885. https://doi.org/10.1093/jssam/smab035

Kuhn, P., & Mansour, H. (2014). Is internet job search still ineffective? *The Economic Journal, 124*(581), 1213–1233. https://doi.org/10.1111/ecoj.12119

Lambert, A. D., & Miller, A. L. (2015). Living with smartphones: Does completion device affect survey responses? *Research in Higher Education, 56*, 166–177. https://doi.org/10.1007/s11162-014-9354-7

Lamas, C. (2005). *The value of coincidental surveys to monitor the validity of tv meter panel measurements* [Paper presentation]. EMRO Conference, Kwa Maritane, South Africa. https://www.aimc.es/a1mc-c0nt3nt/uploads/2010/10/emro2005.pdf

Link, M. W., Murphy, J., Schober, M. F., Buskirk, T. D., Hunter Childs, J., & Langer Tesfaye, C. (2014). Mobile technologies for conducting, augmenting and potentially replacing surveys: Executive summary of the AAPOR task force on emerging technologies in public opinion research. *Public Opinion Quarterly, 78*(4), 779–787. https://doi.org/10.1093/poq/nfu054

Mansouri, B., Zahedi, M., Campos, R., & Farhoodi, M. (2018). Online job search: Study of users' search behavior using search engine query. *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*. Association for Computing Machinery, 1185–1188. https://doi.org/10.1145/3209978.3210125

Ochoa, C., Cortina, C., & González, M. J. (2023, June, 27–29). *Gender differences in job application requirements: Do women demand more of themselves than men? A survey-based study of job-seeking behavior in Spain* [Conference presentation]. 29th International Conference of Europeanists, Reykiavik, Iceland. https://www.upf.edu/documents/244683118/246905697/CEC_2023+%281%29.pptx

Ochoa, C., & Porcar, J. M. (2018). Modeling the effect of quota sampling on online fieldwork efficiency: An analysis of the connection between uncertainty and sample usage. *International Journal of Market Research, 60*(5), 484–501. https://doi.org/10.1177/1470785318779545

Ochoa, C., & Revilla, M. (2022). Acceptance and coverage of fast invitation methods to in-the-moment surveys. *International Journal of Market Research, 64*(5), 565–574. https://doi.org/10.1177/14707853221085204

Ochoa, C., & Revilla, M. (2023). Willingness to participate in in-the-moment surveys triggered by online behaviors. *Behavior Research Methods, 55*, 127–1291. https://doi.org/10.3758/s13428-022-01872-x

Revilla, M. (2017). Analyzing survey characteristics, participation, and evaluation across 186 surveys in an online opt-in panel in Spain. *methods, data, analyses, 11*(2), 135–162. https://doi.org/10.12758/mda.2017.02

Revilla, M. (2022). How to enhance web survey data using metered, geolocation, visual and voice data? *Survey Research Methods, 16*(1). https://doi.org/10.18148/srm/2022.v16i1.8013

Revilla, M., Couper, M. P., Paura, E., & Ochoa, C. (2021). Willingness to participate in a metered online panel. *Field Methods, 33*(2), 202–216. https://doi.org/10.1177/1525822X20983986

Revilla, M., & Ochoa, C. (2015). What are the links in a web survey among response time, quality, and auto-evaluation of the efforts done? *Social Science Computer Review, 33*(1), 97–114. https://doi.org/10.1177/0894439314531214

Revilla, M., Ochoa, C., Iglesias, P., & Antón, D. (2022). *WebdataNow: A tool to send in-the-moment surveys triggered by passive data*. OSF. http://doi.org/10.17605/OSF.IO/G3MSC

Revilla, M., Paura, E., & Ochoa, C. (2021). Use of a research app in an online opt-in panel: The Netquest case. *Methodological Innovations, 14*(1). https://doi.org/10.1177/2059799120985373

Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103*(4), 734–760. https://doi.org/10.1037/0033-295X.103.4.734

Shiffman, S., Stone, A. A., & Hufford, M. R. (2008). Ecological momentary assessment. *Annual Review of Clinical Psychology, 4*, 1–32. https://doi.org/10.1146/annurev.clinpsy.3.022806.091415

Smith, A. (2015). *Searching for work in the digital era*. Pew Research Center. https://www.pewresearch.org/internet/2015/11/19/searching-for-work-in-the-digital-era

Tourangeau, R. (2000). Remembering what happened: Memory errors and survey reports. In A. A. Stone, J. S. Turkkan, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The science of self-report: Implications for research and practice* (pp. 29–47). Lawrence Erlbaum Associates Publishers.

van Berkel, N., Ferreira, D., & Kostakos, V. (2017). The experience sampling method on mobile devices. *ACM Computing Surveys, 50*(6), Article 93, 1–40. https://doi.org/10.1145/3123988

Walker, W. R., & Skowronski, J. J. (2009). The Fading affect bias: But what the hell is it for? *Applied Cognitive Psychology, 23*(8), 1122–1136. https://doi.org/10.1002/acp.1614

Appendix 1. Distribution of Control Variables

Table A1 presents the distribution of the control variables used in the regression analyses for each group.

Table A1 Distribution of control variables in in-the-moment (ITM) and conventional (Conv) surveys

	ITM	Conv
Age (mean)*	41.7	38.6
Gender		
Male	44.6%	38.8%
Female	55.4%	61.2%
Education		
Low-Mid	55.4%	55.2%
High	44.6%	44.8%
Survey participations – last 3 months (mean)*	33.1	27.2
Metered*	100.0%	27.9%
Survey device*		
PC	53.1%	28.4%
Mobile (smartphone and tablet)	46.9%	71.6%

Notes: *indicates a significant effect (5% level) between ITM and Conv.

Appendix 2. List of Job Search Websites

Table A2 lists the job search websites that triggered in-the-moment surveys and the regular expressions used to identify the URLs shown to applicants.

Table A2 List of job search websites

Website	Regular expression identifying a job application
infojobs.net/	infojobs\.net\/candidate\/application\/apply
ticjob.es/	ticjob\.es\/\S*submit-application ticjob.es\/esp\/\S*?status=applied ticjob.es\/esp\/\S*?applied
es.indeed.com/	indeed\.com\/\S*\/post-apply es\.indeed\.com\/pagead\/clk
es.jooble.org/	es.jooble.org\/away\/
infoempleo.com/	infoempleo\.com\/killerquestion\/ infoempleo\.com\/inscription
Tecnoempleo.com	tecnoempleo\.com\/\S*enviar\.php
monster.com	www\.monster\.es\/\S*apply
Randstad.es	randstad\.es\/\S*\/apply\/ randstad\.es\/candidatos\/ofertas-empleo\/\S*\/gracias
Adecco	4dec\.co\/\S*applyFinishOK
Trabajos.com	trabajos\.com\/\S*\/oferta-respondida
Jobatus.es	jobatus\.es\/oferta-trabajo\/\S*?jc=True

Appendix 3. List of Indicators

The following tables summarize the quality indicators, substantive measures used to assess survey differences, and the variables for estimating these indicators.

Table A3.1 Survey evaluation questions

Variables	Question wording and recoded categories
Effort	To what extent did you find it easy or difficult to respond to this survey? ☐ Very easy (recoded as “easy”) ☐ Quite easy (recoded as “easy”) ☐ Neither easy nor difficult ☐ Quite difficult ☐ Very difficult
Satisfaction	To what extent did you like or dislike responding to this survey? ☐ I liked it a lot (recoded as “high”) ☐ I liked it quite a bit (recoded as “high”) ☐ I neither liked nor disliked it ☐ I disliked it quite a bit ☐ I disliked it a lot
Privacy	To what extent do you trust or distrust that your responses to this survey are truly anonymous? ☐ I trust completely (recoded as “trust”) ☐ I trust quite a bit (recoded as “trust”) ☐ I neither trust nor distrust ☐ I distrust quite a bit ☐ I distrust completely
Intrusiveness	To what extent did you find this survey intrusive or not? ☐ Totally intrusive (recoded as “intrusive”) ☐ Very intrusive (recoded as “intrusive”) ☐ Moderately intrusive ☐ Slightly intrusive ☐ Not at all intrusive
Do it again	Would you participate in a survey like this again? ☐ Yes ☐ No ☐ Not sure [if gender = female]

Table A3.2 Quality indicators – Non-recall

Variables	Type	Calculated as …
☐ Company name ☐ Job description	Open-ended	% of answers declaring not remembering or giving non-specific answers
Information in the job description: ☐ Salary ☐ Type of contract ☐ Required experience ☐ Offered perks	Single-response	% of "don't remember" answers
☐ Percentage of met requirements ☐ Percentage of job features that did not fit expectations In the last 6 months: ☐ Number of applications ☐ Number of applications without meeting requirements ☐ Number of applications made without fitting expectations	Numerical open-ended	% of "don't know/remember" answers
☐ Specific not-met requirements	Multiple-response	% of "don't know/remember" answers
List of non-fitting job features: ☐ Salary ☐ Hours ☐ Flexibility ☐ Location ☐ Tasks ☐ Manager ☐ Company ☐ Contract	Set of single-response questions	% of "don't remember" answers
☐ Job search website ☐ Last application date	In-the-moment: Metered data Conventional: Open-ended	In-the-moment: 100% informed Conventional: % of "don't know" answers

Table A3.3 Quality indicators – Invalid answers

Variables	Type	Calculated as …
☐ Company name ☐ Job description Information in the job description: ☐ Salary (specify which) ☐ Type of contract (specify which) ☐ Required experience (specify which) ☐ Offered perks (specify which) Reasons for ☐ Applying without meeting requirements ☐ Applying without a perfect fit	Open-ended	% of invalid answers (not answering what was asked)

Table A3.4 Quality indicators – Length of answers

Variables	Type	Calculated as …
Reasons for ☐ Applying without meeting requirements ☐ Applying without a perfect fit	Open-ended	Mean number of characters

Table A3.5 Quality indicators – Straight-lining

Variables	Type	Calculated as …
☐ Job details (4 questions sharing the same three answer categories) ☐ Fit of features (8 questions sharing the same four answer categories)	Set of single-response	% of respondents selecting the same answer option in all the questions withing the set

Table A3.6 Quality indicators – Inconsistencies

Variables	Type	Calculated as …
☐ Percentage of met requirements ☐ Percentage of job features that did not fit expectations ☐ Probability of being interviewed ☐ Probability of being hired	Numerical open-ended	% of answers outside the range of 0-100%
☐ Percentage of met requirements (< 100%) & requirements not met (= none) In the last 6 months: ☐ Number of applications without meeting requirements < number of applications ☐ Number of applications without a perfect fit < number of applications	Numerical open-ended (except requirements not met, that is multiple-response)	% of combined answers

Table A3.7 Substantive indicators

Variables	Type	Calculated as …
☐ Percentage of met requirements ☐ Proportion of non-compliant participants (applying without meeting all requirements) ☐ Percentage of job features that did not fit applicant's expectations ☐ Proportion of non-fitting participants (applying without a perfect fit) ☐ Probability of being interviewed ☐ Probability of being hired In the last six months ☐ Proportion of non-compliant participants (applying without meeting all requirements) ☐ Proportion of non-fitting participants (applying without a perfect fit)	Numerical open-ended	Differences in means or proportions
Control variables: ☐ Conformity (average score of 11 questions using 5-point scales) ☐ Efficacy (average score of eight questions using 5-point scales)	Sets of 5-point scale questions	Differences in means

Reflective Appendix

This appendix examines the methodological challenges during the experiment design and setup, the unforeseen issues encountered during setup and data collection, and the strategies used to address them.

Foreseen Challenges

In this project, we anticipated several methodological and practical challenges prior to fielding, some of which required adaptations to our original plan. These challenges primarily included (1) the limited representativeness of the metered panel, (2) difficulties in detecting job applications on certain webpages that do not provide unique URLs for such events, (3) the inability to detect in-app job applications on iOS or Android operating systems, (4) challenges in customizing in-the-moment surveys with specific job offer details, (5) difficulties in comparing survey responses with actual job data, and (6) the need to adapt questionnaires for in-the-moment administration by adding screening questions to protect the private information of non-panelists who might use the panelists' metered device to apply for a job.

These limitations were acknowledged and addressed in the main paper. Since these challenges were effectively managed using established strategies, this appendix focuses on the unforeseen challenges encountered during the project and the measures implemented to address them.

Unforeseen Challenges

Setup

The setup phase revealed new limitations of both the software and the method itself:

1. *Non-identifiable URLs:* Some websites did not display specific URLs when applying for a job, making these events undistinguishable from others. Additionally, some websites redirected to employers' sites without showing an identifiable URL for the event of interest. Consequently, four websites (linkedin.com/jobs, jobtoday.com, insertia.net, primerempleo.com) had to be excluded, reducing the ability to detect job applications. In other cases, related events (e.g., initiating the job application process) were used to trigger the survey rather than the actual job application event. In these cases, participants who did not progress to the event of interest were discarded in the questionnaire.

2. *Triggering URLs may change over time:* Websites deploying new versions resulted in changes to triggering URLs, necessitating a monthly repetition of the URL identification process that had not been initially planned.

3. *Job offer identification:* URLs displayed during job applications often lost reference to the job offer, preventing the planned prepopulation of surveys with specific job details such as company names, although job site and time could still be used. This also hindered the direct comparison between survey responses and job offer details that was initially planned, limiting the ability to assess the validity of the survey answers.

A contingency plan to record all web content viewed by participants was considered but not implemented due to the high sensitivity of the collected data and the need for additional approvals. This approach may be explored in a follow-up study.

Fieldwork Execution

In-the-moment fieldwork execution was expected to be much slower than for conventional surveys and affected by the seasonality of the events of interest. The reduced ability to detect applications, as discussed in the "Foreseen challenges" section, combined with the decrease in job applications during July and August due to the vacation period in Spain, required an extension of fieldwork into September to compensate.

Practical Recommendations and Future Developments

In addition to the study's conclusions, researchers working with in-the-moment surveys and similar data types should consider the following recommendations:

1. Dedicate resources to the technology: In this project, enhancing the integration between in-the-moment surveys and metered data allowed us to assess the actual effect of time on participant's responses. In general, effective use of new data types requires specialized technology or careful revisions to existing technologies.

2. Address metered data errors: Researchers should recognize and address errors in metered data (often overlooked), such as participants using non-metered devices or the non-detection of app events, as these issues can impact new methods built on such data and affect feasibility. The Total Error framework for digital traces (TEM) by Bosch and Revilla (2022) provides a comprehensive description of these errors.

3. Embrace technology and internet knowledge: Researchers should have a solid understanding of web and app technologies, as well as internet pro-

tocols. This knowledge is essential for making well-informed decisions and overcoming unexpected challenges during the project, such as realizing that websites with non-identifiable URLs had to be discarded, as was the case in this project.

4. Evaluate pros and cons: Assess when in-the-moment surveys are beneficial. This project illustrates the challenges faced: while they can provide better data than conventional surveys, in-the-moment surveys are time-consuming, require significant support (especially technological), and often result in smaller sample sizes. Therefore, careful feasibility assessment is crucial.

Finally, future research using metered data could benefit from two improvements not available during this project. First, the ability to detect events within apps, which has been recently added to the current version of the meter used. Second, increased coverage of panelists sharing multiple devices. These improvements should reduce false negatives, expanding the sample and/or shortening fieldwork times.