Behaviour & Information Technology

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/tbit20

Fair compensation of crowdsourcing work: the problem of flat rates

Joni Salminen, Ahmed Mohamed Sayed Kamel, Soon-Gyo Jung, Mekhail Mustak & Bernard J. Jansen

To cite this article: Joni Salminen, Ahmed Mohamed Sayed Kamel, Soon-Gyo Jung, Mekhail Mustak & Bernard J. Jansen (2022): Fair compensation of crowdsourcing work: the problem of flat rates, Behaviour & Information Technology, DOI: 10.1080/0144929X.2022.2150564

To link to this article: https://doi.org/10.1080/0144929X.2022.2150564

Published online: 28 Nov 2022.

BEHAVIOUR & INFORMATION TECHNOLOGY
https://doi.org/10.1080/0144929X.2022.2150564

Fair compensation of crowdsourcing work: the problem of flat rates

Joni Salminen[a], Ahmed Mohamed Sayed Kamel [b], Soon-Gyo Jung[c], Mekhail Mustak [d] and Bernard J. Jansen[e]

[a]Department of Marketing and Communication, University of Vaasa, Vaasa, Finland; [b]Department of Clinical Pharmacy, Cairo University, Cairo, Egypt; [c]Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar; [d]Department of Marketing and Sales Management, IESEG School of Management, Lille, France; [e]Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar

ABSTRACT

Compensating crowdworkers for their research participation often entails paying a flat rate to all participants, regardless of the amount of time they spend on the task or skill level. If the actual time required varies considerably between workers, flat rates may yield unfair compensation. To study this matter, we analyzed three survey studies with varying complexity. Based on the United Kingdom minimum wage and actual task completion times, we found that more than 3 in 4 (76.5%) of the crowdworkers studied were paid more than the intended hourly wage, and around one in four (23.5%) was paid less than the intended hourly wage when using a flat rate compensation model based on estimated completion time. The results indicate that the popular flat rate model falls short as a form of equitable remuneration, when perceiving fairness in the form of compensating one's time. Flat rate compensation would not be problematic if the workers' completion times were similar, but this is not the case in reality, as skills and motivation can vary. To overcome this problem, the study proposes three alternative compensation models: *Compensation by Normal Distribution*, *Multi-Objective Fairness*, and *Post-Hoc Bonuses*.

ARTICLE HISTORY
Received 27 October 2021
Accepted 23 October 2022

KEYWORDS
Crowdsourcing; online data collection; fairness; compensation; research ethics

1. Introduction

Crowdsourcing (part of what is referred to as the 'gig economy', Sutherland et al. 2020) is used to collect data for research purposes, such as studying social science phenomena via surveys and questionnaires (Davidson et al. 2017; Salminen et al. 2018a) or annotating machine learning datasets (Alonso 2015; Alonso, Marshall, and Najork 2013; Alonso, Rose, and Stewart 2008). Given the proliferation of remote data collection (partly due to the Covid-19 pandemic making in-person studies more difficult or even impossible), researchers increasingly rely on crowdsourced data collection for obtaining insights about specific subpopulations, or to complete human intelligence tasks (HITs) (Hanrahan et al. 2021). Consequently, the issue of equitable compensation for crowdworkers has become an important topic (Silberman et al. 2018). However, determining an 'appropriate' level of remuneration ('price' of labour, in economic terms) is extremely multifaceted, involving aspects such as the quality of work outputs (Alonso 2015; Alonso, Rose, and Stewart 2008; Kittur, Chi, and Suh 2008), the time taken to complete the work (Cheng, Teevan, and Bernstein 2015; Saito et al. 2019; Whiting, Hugh, and Bernstein 2019), the level of crowdworker expertise (Hanrahan et al. 2021), and the local standards of living (Morrell, Griffin, and Nelson 2018; Narula et al. 2011).

Even though the pricing of crowdwork has been studied from multiple angles, there is a generally agreed premise that the price should be set in a way that is perceived to be 'fair' by the researchers using crowdsourcing to collect data, the crowdworkers who spend their time producing the data, and the platforms mediating crowdwork – in other words, the key parties in this exchange of labour for monetary compensation (Sutherland et al. 2020; Toxtli, Suri, and Savage 2021). Determining fair compensation is not straightforward, however. One prominent issue is that the notion of fairness can have multiple definitions. For example, a flat rate per work output could be considered fair, based on the notion that 'it is the result that matters'. We refer to this conception of fairness as *performance-driven fairness*. An alternative view is that a crowdworker should always be compensated based on their working time, which is referred to as *time-based fairness*. The distinction between these two conceptions (i.e. time versus skills) is crucial, because, for some crowdworkers, the earnings from their work constitute a major source of income (Hara et al. 2018; Savage et al. 2020; Silberman et al. 2018). Consequently, time is of the essence for

CONTACT Mekhail Mustak m.mustak@ieseg.fr 3 Rue de la Digue, 59000 Lille, France

Table 1. Strengths and weaknesses of various crowdworker compensation models.

Model	Strengths	Weaknesses
Fixed flat rate	• Default option in many platforms (i.e. easy to deploy for researchers) • Easy to estimate the total cost of data collection • If all workers have equal cognitive aptitude, skills and speed, there should be no fairness issue	• Workers may differ by their cognitive aptitude, skills and speed, potentially placing some of them at a disadvantage • Ignores the quality of outputs (e.g. an expert worker can carry out the task faster with higher quality than a low-skill worker) • Workers in different locations face different cost of living
Time-adjusted model	• Enables all workers to be compensated based on a generally acceptable hourly wage	• Ignores the quality of outputs (e.g. an expert worker can carry out the task faster with higher quality than a low-skill worker) • Workers in different locations face different costs of living
Quality-adjusted model	• Gives workers an incentive to focus and do well	• Measuring quality is not straightforward and sometimes not even possible (e.g. subjective tasks or surveys asking about opinions) • Workers in different locations face different costs of living
Cost of living adjusted model	• Considers natural wage differences among countries	• Difficulties in obtaining true physical location (e.g. workers from low-income countries might use VPN to pretend to be from a high-income country) • Does not consider the quality of outputs • Does not consider the variation in task completion times

them, and they require a certain minimum hourly compensation in order to make a decent living.

Therefore, for researchers, there is no universally accepted definition of what constitutes 'fair compensation' in crowdsourcing (Singer and Mittal 2013). Perceptions may vary among both researchers and crowdworkers as to what constitutes 'fair' (Whiting, Hugh, and Bernstein 2019). However, there are some principles that might direct what *could* constitute a fair compensation (see Table 1 for their strengths and weaknesses):

- A fixed flat rate, regardless of time spent per unit of output; this would be a performance-based model of crowdwork compensation, also referred to as the *flat rate model*. The rationale is that the same amount of work commands the same level of compensation, regardless of other factors at play.
- *Time-adjusted model*; the workers are compensated based on the time taken to complete the task. Hence, the workers that spend more time (perhaps because they are less skilled) end up receiving a higher compensation than those completing the task faster. The rationale is that one should compensate workers so that the overall time they invest in carrying out tasks results in an acceptable hourly wage.
- *Quality-adjusted model*; the workers that produce higher quality outputs are compensated more than those producing lower quality outputs. The rationale is that better workers deserve more pay.
- *Cost of living adjusted model*; in which the crowdworker is compensated based on their physical location (e.g. a worker in India would be paid substantially less than a worker in the UK because the latter has a much higher cost of living). The rationale is that the price of labour naturally varies by country, which should be reflected in the level of compensation.

Of these four models, the most commonly used in research practice is the flat rate model, primarily due to its convenience and the historical tradition of compensating research participation on the basis of output. However, even in the case of a flat rate, researchers often attempt to factor in how demanding the task is, i.e. how much time it takes for an average worker to complete it. This estimation is typically done via pilot testing. Factoring in time based on pilot testing (Gaikwad et al. 2015; Silberman et al. 2018) has the appeal of simplicity, where all a researcher needs to know is an estimation of the average time taken to complete the task and the hourly minimum wage of a given location. This yields a simple calculation: $reward = time\ taken\ in\ hours \times hourly\ rate$. However, leaving aside economic discussions of the fairness of minimum wages (Stigler 1946), the problem with this approach is that it is based on the assumption of fixed (or predictable) completion times.

In reality, the time taken to finish a task may vary greatly among crowdworkers. If the variation is considerable, then a number of participants can be paid either more or less than the hourly wage intended by the researcher, which, if one subscribes to the notion of time-based fairness, can leave some workers disadvantaged while using budgeted funds for workers that work faster. Faster work, in turn, can be achieved from many standpoints, for example, a higher cognitive capacity, better expertise, or strategising approaches to answer as quickly as possible (i.e. 'speedrunning'). Symmetrically, crowdworkers that would find it difficult to keep up with generally fast completion times, easily risk being under-compensated relative to the time they put into the task, because faster workers would tilt the average completion time (based on which the flat rate is calculated) to a lower level than that required for slower workers to be fairly compensated.

Overall, the problem of fair compensation in crowdsourcing is far from being resolved, with several challenges still lingering. The issue is curbed by at least two additional considerations: (1) there is an uneven distribution of power among crowdworkers and researchers, as the latter typically set the prices unilaterally (Gaikwad et al. 2015); (2) many of the present crowdsourcing platforms do not provide adequate tools for supporting fair compensation models other than the flat rate model. Therefore, there is a practical design problem where features and functionalities within crowdsourcing platforms do not readily permit fair compensation by either default or out-of-the-box means (Fox et al. 2020), and fair compensation is left to the discretion of researchers. If treated lightly or without due consideration, researchers may inadvertently contribute to creating working conditions that yield unethical outcomes. In the extreme, a systematic disregard for crowdworkers' rights could result in the creation of a 'new global underclass' (Gray and Suri 2019, 1).

Against this backdrop, the aim of this study is to empirically investigate the fair compensation problem. In this context, we also investigate the accuracy of researcher-estimated completion times and whether there are demographic factors that influence how long it takes a worker to finish a task. Our research questions (RQs) are:

- **RQ1:** *Do completion times in crowd studies vary by participants?* By addressing this question, we can determine the validity of the time constant method for determining fair remuneration. We define 'time constant' as a fixed payment based on the estimated, not actual, time required to finish a survey by a crowdworker.
- **RQ2:** (a) *How many participants go over the estimated completion time for each study?* (b) *How many go under the estimated time?* Addressing this question will inform us about the 'importance' of time variability in terms of appropriate remuneration.
- **RQ3:** *Do the actual completion times match the estimated completion times in crowd studies?* Addressing this question will reveal how 'tough' or 'accurate' it is to estimate completion timeframes in crowdsourced data collecting.
- **RQ4:** *Are there demographic factors (age, gender, nationality, education, etc.) that correlate with the completion times in crowd studies?* Addressing this question will help us detect possible patterns in crowdworkers' natural ability to complete tasks quickly or slowly.
  - **RQ4a.** *Are there* demographic *factors that correlate with a participant going over the estimated completion time?*
  - **RQ4b.** *Are there* demographic *factors that correlate with a participant going under the estimated completion time?*
  - **RQ4c.** *Are there* demographic *factors that separate 'speed-runners' from 'slow-runners'?* We define speed-runners as crowdworkers who complete the task in substantially less time than estimated (standard deviations (SD) –3), and slow runners as crowdworkers who complete the task in substantially more time than estimated (SD +3).
- **RQ5:** *How many participants are under- or over-compensated relative to minimum wage?* Addressing this question will help us determine, in monetary terms, what the ramifications of time variability for fair compensation are.
- **RQ6:** *Do researchers and workers win or lose in monetary terms by setting a fixed price?* This question helps us address how the parties 'win' or 'lose' based on a constant estimate and actual time variability, when the compensation is fixed at a flat rate.

Our study addresses a critical issue faced by researchers using crowdsourcing for survey research. Fairness is critical not only for 'doing the right thing', but also because equitable compensation is likely to result in increased worker motivation and, thus, higher quality results (Kaufmann, Schulze, and Veit 2011; Wang, Yang, and Wang 2019). To this end, we contribute to the emerging field of the ethical examination of crowdworking conditions by analyzing crowdworkers' compensation fairness on the basis of hourly wage. Additionally, we identify and articulate challenges for estimating task completion times for research. Here, we focus on the viewpoints of the workers and researchers, while leaving the viewpoint of the platforms for future research.

2. Related work

We define fairness as *a level of payment that corresponds to a commonly accepted hourly wage* – i.e. a fair payment for a given worker is one that: (a) equals or exceeds a standard wage for the job (e.g. a minimum hourly wage) based on the actual time the worker contributes to the task; but (b) does not significantly exceed this standard. The first part of this definition incorporates the workers' perspective, and the second part incorporates the researchers' perspective, although we discuss the nuances of this definition in Section 5. We adopt the concept of time-based fairness as a choice of believing that crowdworkers' compensation should be in line with the time they take for completing a task. Again, as mentioned in the previous section, there are multiple definitions for fairness, and our choice represents one of those definitions. Here, we review the related work in this field.

Crowdsourcing platforms provide online users with a centralised marketplace to earn money, in exchange for their time and effort (Toxtli, Suri, and Savage 2021). Setting the 'right price' is one of the fundamental activities in crowdsourcing (Cheng, Teevan, and Bernstein 2015). Researchers (also known as 'requesters') can set too low prices to satisfy workers' demands, but can also set too high prices for what would be considered as a fair 'market price' for a given task. Studies in this field can be broadly divided into two categories: (a) those dealing with what constitutes just, fair, and right compensation for crowdworkers' time (Barbosa and Chen 2019; Irani and Six Silberman 2013; Whiting, Hugh, and Bernstein 2019; Silberman et al. 2018), and (b) those focused on analyzing pricing strategies, behaviours, and optimal (efficient) compensation from an economic point of view (Hong and Pavlou 2012; Mao et al. 2013a; Sheng, Provost, and Ipeirotis 2008; Singer and Mittal 2013; Tong et al. 2018; Wang, Yang, and Wang 2019). It is generally acknowledged that these two viewpoints may not always coincide, and a researcher may pit crowdworkers to compete against one another in a 'race to the bottom' (Jäätmaa 2010) by gradually decreasing the price until he or she finds the cheapest worker satisfying the given quality criteria. Nonetheless, viewpoints of both justice and economic rationality are likely to be needed to address the compensation problem in crowdsourcing.

In another line of work, underpayment has been reported as the most impactful concern that workers face on Amazon Mechanical Turk (MTurk) (Whiting, Hugh, and Bernstein 2019). A large-scale empirical analysis estimates that the effective hourly wages among workers in MTurk ('turkers') are approximately \$2 to \$3 USD (Hara et al. 2018), which is less than half of the federal minimum wage (\$7.25 USD per hour) in the United States (US) even though three-quarters of the active turkers reside in the US (Difallah, Filatova, and Ipeirotis 2018). The study further shows that only 4% of turkers earn more than the federal minimum wage (Hara et al. 2018). Paradoxically, a study by Whiting, Hugh, and Bernstein (2019) found that workers considered \$15 as a fair hourly compensation. This means that to meet the workers' expectations, the crowdsourcing industry needs to quadruple the pay. Hence, it is appropriate to argue that (at least some) 'workers in crowd markets struggle to earn a living' (Saito et al. 2019, 3187) and face issues such as unpaid labour, and a lack of healthcare or social security benefits (Gray and Suri 2019; Toxtli, Suri, and Savage 2021). Some academics have issued dire warnings of an emerging 'global underclass', where humans work for the minimum wage to provide inputs for algorithms; entities that are perceived to hold more power than the expendable worker (Gray and Suri 2019). Irani and Six Silberman (2013) invoke the concept of 'worker invisibility' to emphasise the hidden human labour behind algorithms; that is, humans have contributed their knowledge to develop technological solutions that have been incorporated into dozens of commercial applications, and this labour has frequently occurred with little or no regard for fair compensation (Silberman et al. 2018).

As a solution, Mankar, Shah, and Lease (2017) suggest 'design activism', a form of collective action that aims at improving the conditions of online crowdworkers, based on the notion that crowdworkers are active participants as opposed to mere passive subjects to platform governance. However, crowdworkers may be hesitant to join the activist movement for fear of jeopardising their present platform positions, or simply because they prefer to spend their time earning rather than on what may be viewed as an unpredictable 'fight for change'. This is visible from the findings of Mankar, Shah, and Lease (2017), who indicated only 20% of workers support enforcing a minimum wage requirement. Moreover, Salehi et al. (2015) observed multiple points of friction for crowdworkers to get organised, such as the fragmentation and distance imposed by the online environment. Hence, researchers and platforms have a heightened responsibility for fair compensation in crowdsourcing. This stems from a moral obligation to protect the more vulnerable party in an exchange (Achrol and Gundlach 1999), defined as being one with less bargaining power. In crowdsourcing, workers tend to expose themselves to rejection, mistrust, and risk (McInnis et al. 2016), while also facing a high uncertainty of available jobs (Sutherland et al. 2020). Researchers may reject their work for no valid reason; technical errors may eradicate a half-finished work; the job may be paused before completion; and so on. Because of these challenges, protecting crowdworkers – in the same way that 'real' workers are protected (Erickson et al. 2014; Seidman 2007) – is a crucial undertaking (Irani and Six Silberman 2016).

In practice, attention to this issue has culminated in the maxim of paying crowdworkers *at least the minimum wage* (Silberman et al. 2018). The research community has attempted to develop various tools for supporting the mission of fair compensation. Among these, Whiting, Hugh, and Bernstein (2019) suggested 'Fair Work' (FW), an algorithmic approach used to assess workers to gain an estimate of how long a given task takes and to provide automatic bonuses (where needed) to set the effective wage to correspond with the inputted minimum wage. However, the FW approach focuses on micro-tasks, not surveys (unlike our study). By way of characteristics, surveys and micro-tasks are very different, with the former typically requiring more concentration and effort, and therefore the nature of time-estimation might also be different. Moreover, the FW approach relies on self-reported estimates from the crowdworkers (Whiting, Hugh, and Bernstein 2019), which provides a financial incentive to inflate the estimates and/or to perform the task slower than would otherwise be needed. As our study focuses on actual completion times, its results are less fallible to this bias.

Researchers have created tools to support crowdworkers. For example, Turkopticon (Irani and Six Silberman 2013) enables workers to alert others of consistently low-paying researchers, thereby acting as a reviewing mechanism for increasing trust and potentially eliminating unethical researchers from the platform. The tool exemplifies information sharing as a form of collective action for crowdworkers to maintain satisfactory working conditions. However, as an external tool, it is not integrated into the platforms' design, and can thus be considered as side-information rather than as an integral part of the economic activity taking place on the platform. However, while taking an important step towards addressing knowledge sharing in the crowdsourcing community, the study does not address potential issues emerging from time variability.

Another tool, TurkScanner (Saito et al. 2019), uses machine learning to predict the workers' time to complete a micro-task, thus estimating the expected hourly wage. While this approach can be useful at inferring data-driven estimates for determining a generally fair compensation level, it does not address the same problem that we tackle in this work – namely, the variability in completion times, and how such variability would affect the usefulness of the estimates for a fair pricing of crowd labour. Moreover, these tools may be difficult for researchers to implement. As Whiting, Hugh, and Bernstein (2019) state: '[…] many requesters want to pay fairly, but are put off by the requirement to iterate and prefer a solution that they do not need to babysit' (199). Researchers are tempted to use the default options available in platforms and may perceive that using these platforms to set a fixed price that exceeds the minimum wage based on average completion time would satisfactorily address the problem of fair compensation. But as we illustrate in this study, this is not the case.

Estimating task duration is difficult for both workers and researchers (Saito et al. 2019). However, the consequences for the two parties differ. For workers, underestimating the time taken to complete a task results in working for less than they should be compensated (i.e. a negative outcome). For researchers, underestimating results in a positive outcome that reduces data collection costs and so makes it possible to hire more workers for the same budget. For these reasons, workers have an incentive to work as fast as possible (Mao et al. 2013b), even at the expense of quality, whereas researchers have the incentive of being precarious with setting the compensation level (Slivkins and Vaughan 2014). In an empirical study, Whiting, Hugh, and Bernstein (2019) found that requesters both under- and over-estimated workers' completion times, with under-estimation being a more common tendency and leading to 68% of the requesters underpaying for the work. A particularly important notion here is how workers vary by their natural qualities, such as cognitive capacity, eyesight, typing speed, etc. For example, Hanrahan et al. (2021) argued that group-based differences (e.g. expert and novice workers) could lead to wage imbalances. If workers do vary in their attributes, then a reasonable conceptualisation of fairness considers an equitable representation of workers (Gaikwad et al. 2015) with distinct attributes. In an extreme case, a worker may be slower due to a visual impairment. This worker can complete the task, and has the expertise, but is simply slower due to their physical limitations. However, the currently dominant flat rate paradigm penalises such workers in the sense that they get compensated less for their time than peers without any limitations and/or higher expertise.

3. Methodology

3.1. Datasets

To account for the variability of different study contexts, we investigate the results of three actual user studies (see Table 2). As these are real research studies, one can expect crowdworker behaviours to be realistic and not exhibit some form of 'observer bias' (Mahtani et al. 2018) as might occur with studies on actual crowdworker wages.

Table 2. Primary details of the studies.

	Estimated completion time	Flat rate (£)	Estimated hourly wage (£)	Minimum wage at the time (UK)	Context
Study 1	15 minutes	2.24	8.96	£8.91	Validating a scale
Study 2	15 minutes	2.00	8.00	£8.72	Testing a design
Study 3	25 minutes	3.50	8.40	£8.72	Testing a system

The studies are described in more detail as follows:

- **Study 1** focused on scale validation as a context. The participants were asked questions about their organisation's use of personas. The purpose was to assess the validity and reliability of the developed survey constructs.
- **Study 2** focused on user testing alternative designs of persona profiles as a context. The participants were shown three personas and then asked questions about each. The purpose was to test how manipulation in persona profiles affects user perceptions of personas.
- **Study 3** focused on testing an interactive persona system as a context. The participants were asked to use a persona system and then answer questions about the personas. The purpose was to collect data on users' behaviours and perceptions.

A fixed compensation rate was applied in all studies. The researchers aimed at targeting approximately the United Kingdom (UK) minimum wage at the time of each study. Therefore, each study's estimated hourly wage is approximately equivalent to the minimum hourly wage at the corresponding time (see Table 2). The UK wage was selected as it is among the higher average wages globally – in other words, paying the UK minimum wage has a high chance of satisfying the local minimum wage standards, regardless of the crowdworker's country of origin.

All studies were conducted in Prolific, which is a platform that has been deployed in various user studies and social science research (Kothe and Ling 2019; Palan and Schitter 2018; Peer et al. 2017; Salminen et al. 2019). We chose this platform because it tracks the task completion times, which is a required element we use to address our RQs. The sampling criteria were similar across the three studies (see Table 3). However, the study context contained more variability. We analyzed the results from each study separately, because the studies were conducted at different time points and with different research goals, so their separation is a logical approach. The overall dataset included responses from 363 participants. Study 1 included 123 respondents, while Studies 2 and 3 included 143 and 97 respondents, respectively.

We use the number of steps that the participants had to go through as a proxy for a study's complexity. Previous studies have associated the complexity of task design with the worker's effort and time taken (Cheng, Teevan, and Bernstein 2015), so accounting for the study design matters for interpreting our results. For example, if a participant needs to switch between persona perception assessment and personality assessment (see Study 2 in Table 3), then this is more cognitively labourious (i.e. complex) than having to complete fewer steps. As indicated in Table 3, Study 1 had the lowest complexity (4 steps), followed by Study 3 (7 steps), and then Study 2 (16 steps).

Table 3. Sampling criteria and effort required.

Study 1	Study 2	Study 3
age: 23–62 Nationality: UK, USA, Ireland, Australia, New Zealand Full-time employment Students excluded	age: 25–45 Nationality: UK, USA, Ireland, Australia Full-time or Part-time employment Students excluded	age: 25–64 Industry: Art/Design, Graphic design, Market research Students excluded
Complexity: Low (4 steps)	Complexity: High (16 steps)	Complexity: Medium (7 steps)
current state assessment PRS scale* applicability assessment background information	onboarding persona viewing [P1] first impression assessment [P1] PPS** scale [P1] TIPI*** scale [P1] confidence assessment [P1] persona viewing [P2] first impression assessment [P2] PPS scale [P2] TIPI scale [P2] confidence assessment [P2] persona viewing [P3] first impression assessment [P3] PPS scale [P3] TIPI scale [P3] confidence assessment [P3]	use of system task completion PPS scale trust and transparency scale NASA TLX**** scale technical sophistication scale background information

Notes: * Persona Readiness Scale (Salminen et al. 2021); ** Persona Perception Scale (Salminen et al. 2020); *** Ten Item Personality Measure (Gosling, Rentfrow, and Swann 2003); **** The NASA Task Load Index (Cao et al. 2009).

Table 4. Participant focus of this research.

	Speed	
Quality	Fast	Slow
High	Ethical speed-runners	Ethical slow-runners
Low	Unethical speed-runners	Unethical slow-runners

3.2. Variables

The time-based compensation variables used are:

- **Estimated study completion time**: the time expected to complete the task as determined by the researcher
- **Flat rate**: what the participants were paid, regardless of how long it took them to complete the task
- **Average hourly rate**: the average hourly wage based on actual study completion times
- **Estimated hourly rate**: the average hourly wage based on estimated study completion times
- **Minimum hourly rate**: the minimum hourly wage in the UK at the time of study
- **Actual average completion time**: the average time taken by participants to complete the study

Times were calculated in minutes. The mean ± SD was used to summarise the distribution of the time taken to complete the surveys. The trimmed mean and coefficient of variation (CV) were also calculated for each study. The trimmed mean was calculated after excluding the top and bottom 10% of the data. The CV is a statistical measure of the relative dispersion of data points in a data series around the mean. The higher the coefficient of variation, the greater the level of dispersion around the mean. Conventionally, CV < 10 is considered *excellent*, 10–20 is considered *good*, 20–30 is *acceptable*, and >30 is not *acceptable*. Based on their worker IDs in the crowdsourcing platforms, the participants were prevented from taking part in more than one of the studies.

The time estimates were determined by the researchers based on measuring the time taken by the test participants i.e. the pilot study approach (Thabane et al. 2010) in each study. These formed the actual time estimates that were used in each study. For the actual completion times, the interquartile range (IQR) was defined as the third or upper quartile (Q3) minus the first or lower quartile (Q1). Outliers were defined as points below $Q1 - 1.5 \times IQR$ or above $Q3 + 1.5 \times IQR$. These points were used to define outliers as they are more robust than using mean and SD, and can be used even if the data is not normally distributed. The *z*-score was also used to assess the presence of outliers. Outliers were defined as points above or below the mean of ±3 SD.

The SD that was calculated based on the actual time was used to identify outliers for the estimated time. Outliers were defined as points greater or smaller than the estimated mean of ±3 SD. Finally, we refer to the obtained outliers as ‘speed-runners’ (for those that went significantly under the average or estimated times) and ‘slow-runners’ (for those that went significantly over the average or estimated times). In our operationalisation, the term ‘speedrunner’ does not have a negative connotation, and instead, all selected participants are considered as ethical crowdworkers, as their survey responses passed the attention checks applied in each study. Table 4 illustrates this conceptual division.

As illustrated in Table 4, participants who complete tasks faster (i.e. speed-runners) can conceptually act either ethically (i.e. follow instructions, aim to provide honest quality work) or unethically (not follow instructions, aim only to quickly complete the task without considering the quality of work). Slow-runners can also be either ethical or unethical. In this study, we focus on ethical participants, which is achieved by only including those participants whose work outputs passed the quality control (highlighted in green in Table 4).

4. Results

4.1. RQ1: Do completion times in crowd studies vary by participants?

To address this research question, we examine statistic descriptors of the study completion times (e.g. range, mean, SD), as well as histograms.

As shown in Table 5, the mean completion time varied between studies, and the participant completion time varied for each study, regardless of whether the study task was relatively short (Study 1) or comparatively longer (Study 3). The data for Study 1 were normally distributed, but this was not the case for Studies 2 and 3 which showed skewness and the presence of many outliers (see Figure 1). Although the average time for completion in Study 1 was 5.65 ± 2.23 min, the time for completion varied from a minimum of 2.4 min to a

Table 5. Descriptive statistics for the included studies.

Study	*N*	Mean	SD	Median	Trimmed mean	Min	Max	Range	CV	ET (Minutes)
1	123	5.65	2.23	5.15	5.42	2.4	15.18	12.78	39%	15
2	143	16.48	10.42	12.33	14.82	5.41	55.65	50.24	63%	15
3	97	19.37	11.42	15.29	17.43	6.13	68.99	62.86	59%	25

SD: Standard deviation; CV: Coefficient of variation; MAD: Mean absolute deviation; ET: Estimated time by the researchers.

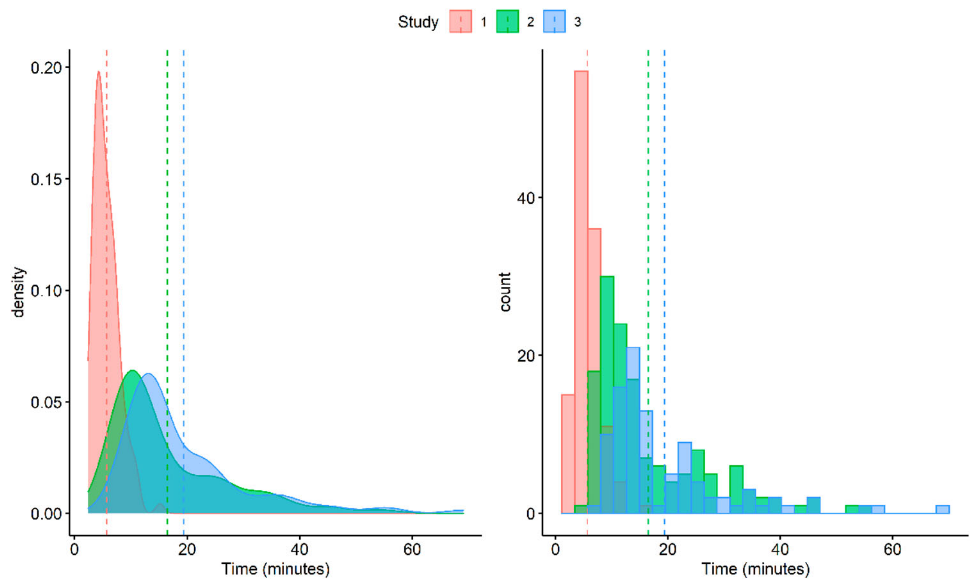

Figure 1. Distribution of time needed to complete the questionnaire.

maximum of 15.18 min. The variability was even greater for Studies 2 and 3, with a range of ∼56 and 62 min, respectively. The CV ranged from 39% to 63% which indicates high variability in the completion time. As such, the answer to RQ1 is that *study completion times vary a great deal by participants; hence, making time-insensitive mechanisms challenging for any fair compensation of crowdworkers at the individual level.*

4.2. RQ2: (a) How many participants go over the estimated completion time for each study? (b) How many go under the estimated time?

To address this research question, we calculate the number of outlier participants that:

(a) *Significantly exceed the estimated completion time.* These participants can be defined as slow-runners relative to the estimated time.

(b) *Significantly go under the estimated completion time.* These participants can be defined as speed-runners relative to the estimated time.

(c) *Significantly exceed the actual mean completion time.* These participants can be defined as slow-runners relative to actual completion time.

(d) Significantly go under the actual completion time. These participants can be defined as speed-runners relative to actual completion time.

We report these separately for each study, and interpret the obtained frequencies.

4.2.1. Study 1

The results for Study 1 (see Figure 2) show that the actual completion time (5.65 ± 2.23) was significantly different from the estimated time of 15 min ($P < 0.001$ using t-test).

Based on the estimated mean, 87.8% of the respondents were identified as speed-runners (see Table 6), as they completed the questionnaire in a time lower than the estimated mean by more than three SD. None were identified as slow-runners. Based on the actual time, only one respondent was identified as a slow-runner (0.81%), while none were identified as speed-runners. Hence, the results from Study 1 indicate a substantially greater portion of speed-runners than slow-runners (see Figure 3) – i.e. the *participants are likely to complete the study faster than expected based on a pilot study, and a sizeable portion of them complete it much faster relative to the overall average completion time.*

4.2.2. Study 2

The results for Study 2 (see Figure 4) show that the average completion time was not significantly different from the estimated time ($P = 0.09$) at the 0.05 level.

Based on the estimated mean, 1.4% of the respondents were identified as slow-runners (see Table 7), as they completed the questionnaire in a time lower than the estimated mean less 3 SD. None were identified as speedrunners. Based on the actual time, five respondents were identified as slow-runners (3.5%) while none were identified as speed-runners, as illustrated in Figure 5. These findings differ from Study 1 in that *Study 1 found substantially more speed-runners than slow-runners, whereas, for Study 2, it is the*

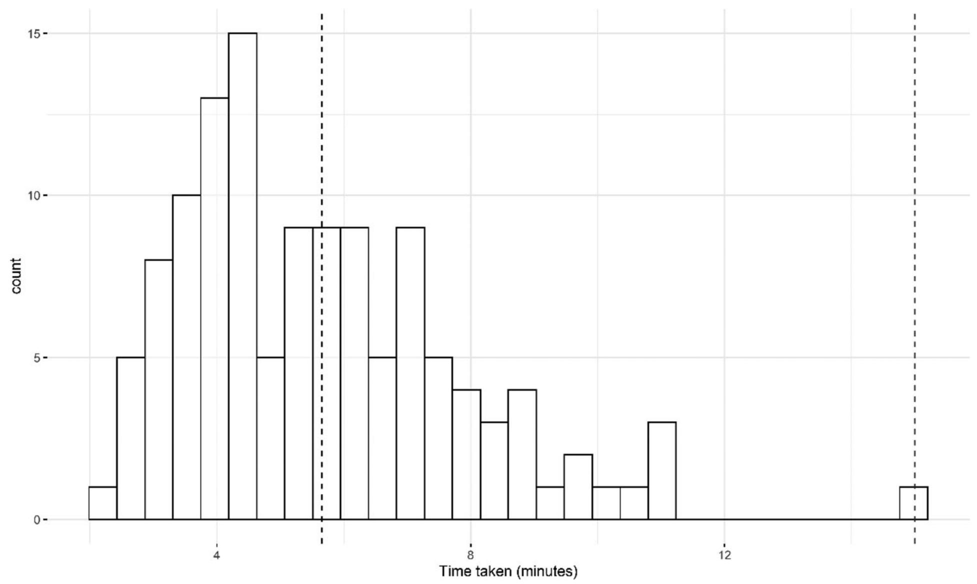

Figure 2. Time taken to complete the questionnaire. The left dotted line represents the actual time while the right dotted line represents the estimated time.

Table 6. Outliers for the estimated and actual time.

Estimated time	Actual time						Estimated time		Actual time	
Mean	Median	Mean	SD	IQR	Q1	Q3	Slow runners	Speed runners	Slow runners	Speed runners
15	5.15	5.65	2.23	2.87	4.1	6.98	0 (0%)	108 (87.8%)	1 (0.81%)	0 (0%)

opposite. However, both studies indicate that the distribution of time taken leans towards less time rather than more time. This is visible in Figures 3 and 5. When comparing Figures 3 and 5, one can see that Study 1 had less bias towards slow-runners, and the density in both figures is biased towards speed-runners. Therefore, Study 2 supports the previous finding that *participants are likely to complete the study faster than expected based on a preliminary pilot study, and a sizeable portion of them complete it much faster relative to the overall average completion time.*

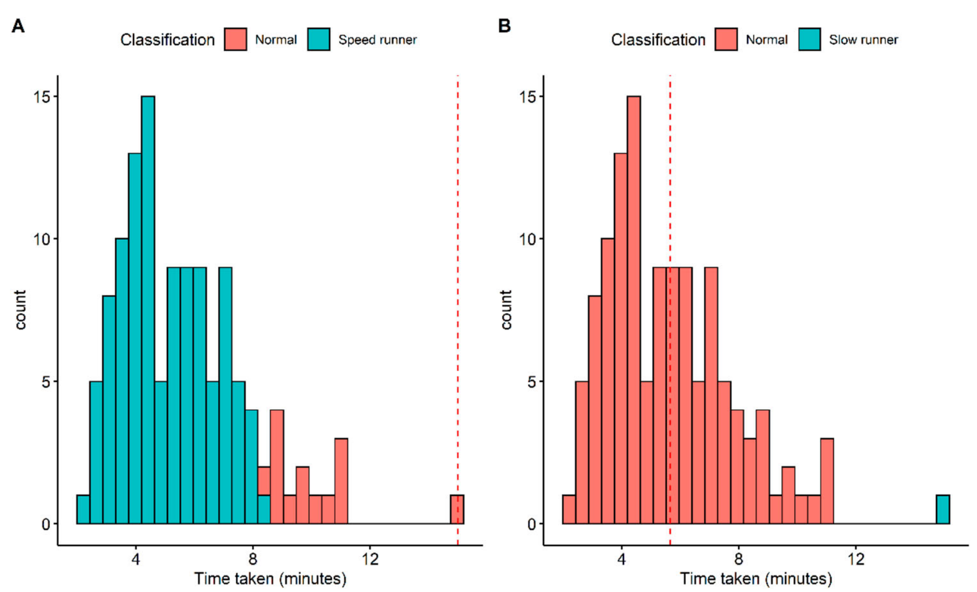

Figure 3. Classification of the respondents based on (A) Estimated time and (B) Actual time.

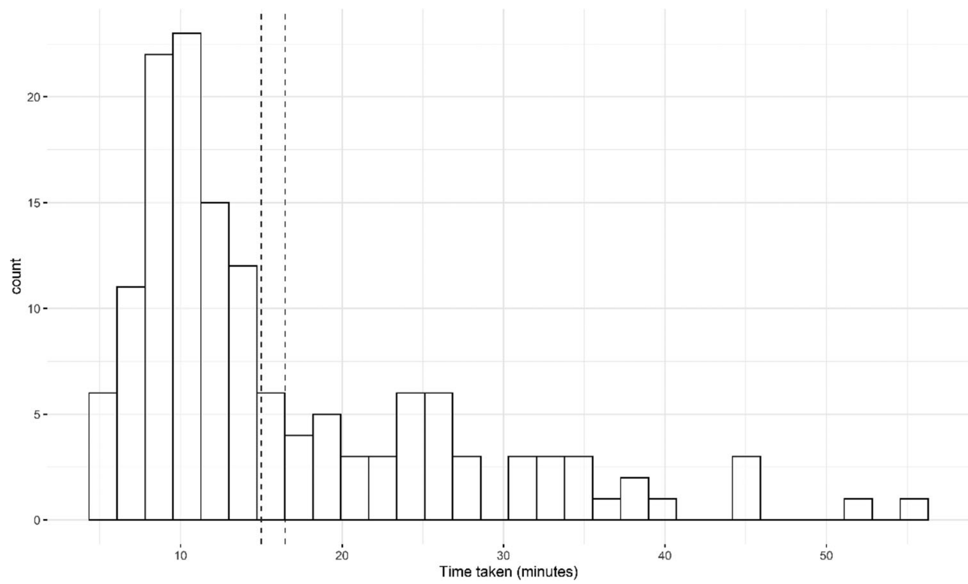

Figure 4. Time taken to complete the questionnaire. The dotted line on the right represents the actual time while the dotted line on the left represents the estimated time.

Table 7. Outliers for the estimated and actual times for completing the questionnaire.

Estimated time	Actual time						Estimated time		Actual time	
Mean	Median	Mean	SD	IQR	Q1	Q3	Slow runners	Speed runners	Slow runners	Speed runners
15	12.33	16.48	10.42	12.2	9.26	21.4	2 (1.4%)	0 (0%)	5 (3.5%)	0 (0%)

4.2.3. Study 3

The results for Study 3 (see Figure 6) show that the average completion time (19.37 ± 11.42) was significantly different from the expected 25 min estimated time ($P < 0.001$ using t-test).

Based on the estimated mean, 1.03% of the respondents were identified as slow-runners (see Table 8), as they completed the questionnaire in a time lower than the estimated mean less 3 SD. None were identified as speed-runners. Based on the actual time, six respondents were identified as slow-runners (6.19%), while none were identified as speed-runners. As previously, the results in Figure 7 support the notion that most participants tend to be fast rather than slow.

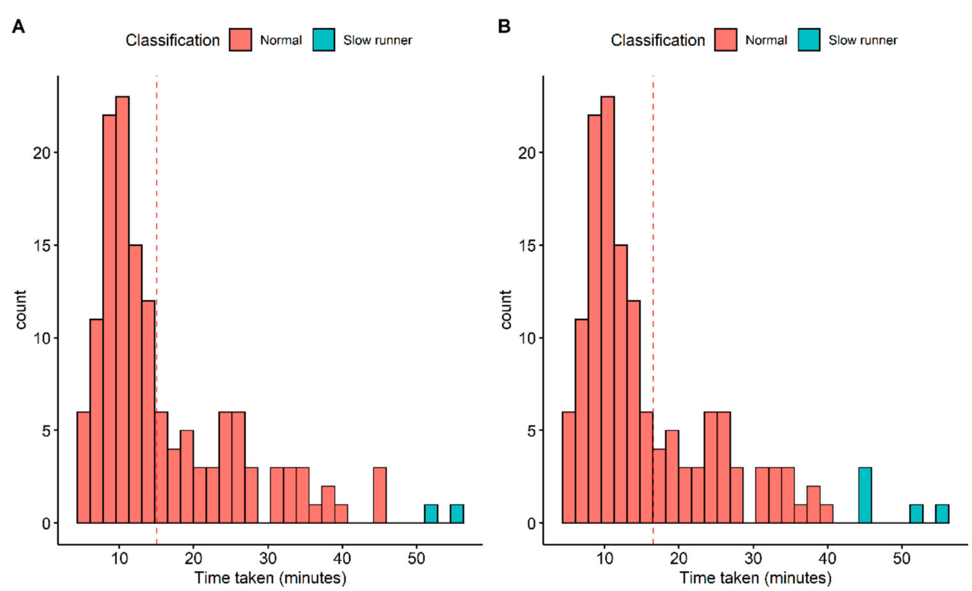

Figure 5. Classification of the respondents based on (A) Estimated time and (B) Actual time.

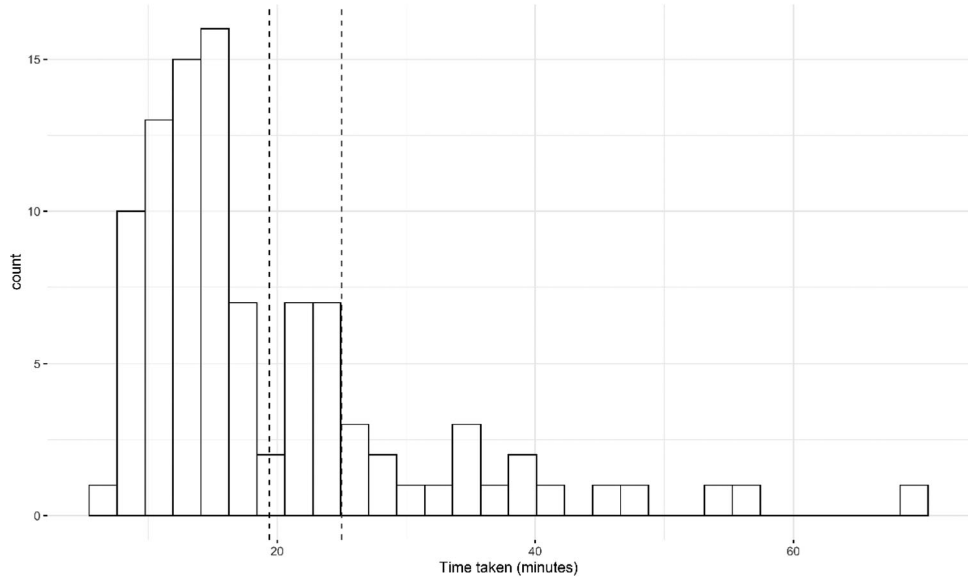

Figure 6. The time taken to complete the questionnaire. The dotted line of the left represents the actual time while the dotted line on the right represents the estimated time.

Table 8. Outliers for the estimated and actual times for completing the questionnaire.

Estimated time	Actual time						Estimated time		Actual time	
Mean	Median	Mean	SD	IQR	Q1	Q3	Slow runners	Speed runners	Slow runners	Speed runners
25	15.29	19.37	11.42	11.24	12.1	23.32	1 (1.03%)	0 (0%)	6 (6.19%)	0 (0%)

4.3. RQ3: Do the actual completion times match the estimated completion times in crowd studies?

As reported in the previous section, in Studies 1 and 3 the estimated and actual completion times were significantly different. However, in Study 2 (the most complex of the studies), the estimated and actual completion times were not significantly different.

4.4. RQ4: Are there demographic factors (age, gender, nationality, education, etc.) that correlate with the completion times in crowd studies?

Linear mixed modelling was used to assess factors associated with the average completion time. The study was included as a random intercept to take into

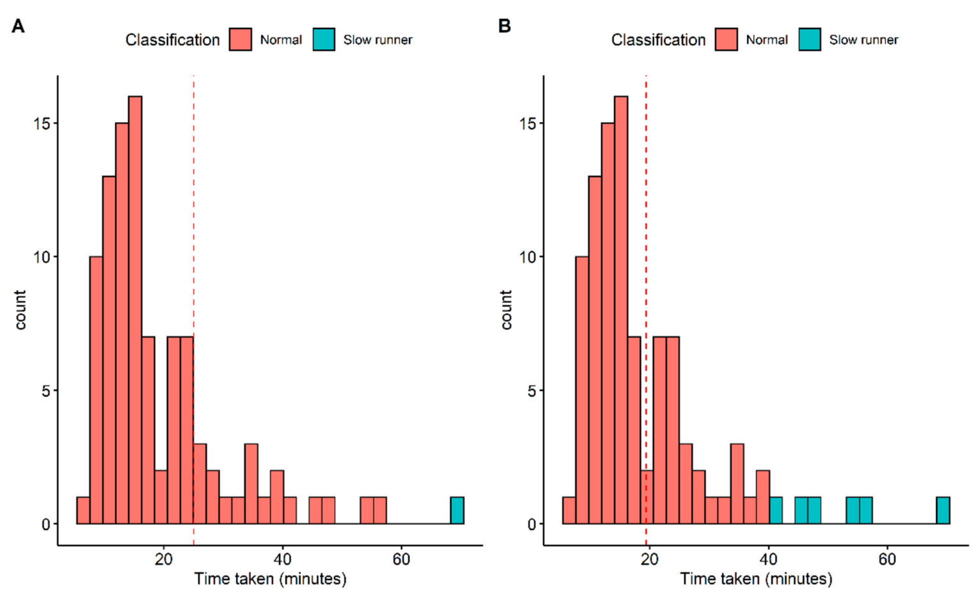

Figure 7. Classification of the participants based on (A) Estimated time and (B) Actual time.

Table 9. Linear mixed modelling analysis results.

	Time taken		
Predictors	Estimates	CI	P
(Intercept)	9.31	0.61–18.02	**0.036**
Age	0.05	−0.05 to 0.16	0.333
Sex [male vs. female]	1.07	−0.68 to 2.82	0.231
Nationality: United Kingdom	Ref		
Nationality [N/A]	16.34	0.17–32.51	**0.048**
Nationality [Other]	3.63	0.83–6.42	**0.011**
Nationality [United States]	4.92	2.58–7.27	**<0.001**
Random effects			
σ^2	66.37		
$\tau_{00\ \text{study}}$	44.32		
ICC	0.40		
N_{study}	3		
Observations		343	
Marginal R^2/Conditional R^2	0.046 / 0.428		

account the varying average completion times between the three studies. Age, gender, and nationality were included as independent variables. We also had some data on the highest education level obtained by the participant, as well as their current employment, but since these variables were missing for most participants, we did not include them in the model. Hypothesis testing was performed at a 5% level of significance.

The results in Table 9 show that neither age ($P = 0.333$) nor gender ($P = 0.231$) was significantly associated with the average time taken to complete the questionnaire. However, nationality showed a statistically significant association with the average time taken to complete the questionnaire. The average time required by respondents from the US (B = 4.92, $P < 0.011$) was significantly higher than the time required by the respondents from the UK. Similarly, the average time required by respondents from other countries was significantly higher than the time required by respondents from the UK (B = 3.63, $P < 0.05$). Addressing RQ4, this indicates that *age or gender are not significant predictors for study completion time, but the country of origin is a significant predictor.* Nationality does correlate with participants going over or under the estimated completion time.

Given that there were only three slow-runners across all studies and most speed-runners were in Study 1, we cannot evaluate RQ04c. However, since age and gender were not significant overall, we can determine that these demographic variables do not impact either going over or under the estimated completion time (RQ04a and RQ04b). Nationality is significant though, with crowdworkers from the UK taking statistically significant less time than workers from other nationalities (RQ04a and RQ04b).

4.5. RQ5: How many participants are under- or over-compensated relative to minimum wage?

To address RQ5, we calculated the actual hourly rate (also referred to as the effective wage, Whiting, Hugh, and Bernstein 2019) for each participant by dividing the assigned reward per completion with the actual time taken. We then compared these rates to minimum UK hourly wages at the time. The flat rate for completing Study 1 was £2.24, while the flat rates for completing Studies 2 and 3 were £2 and £3.50, respectively. The hourly rate was calculated for each participant by dividing the flat fee by the completion time (in minutes), and then multiplying by 60.

The results (see Table 10) show that the calculated actual hourly rate was higher than the minimum hourly wage in all three studies. The actual hourly rate was significantly higher than the estimated hourly rate in all studies ($P < 0.001$ using t-test). The results also showed a higher variability in the actual hourly rate within each study as shown by the CV. Addressing RQ5, across all studies, the flat rate used resulted in an actual hourly rate that was higher than the estimated minimum hourly wage in 76.9% ($n = 279$) respondents. Moreover, the average hourly rate was significantly different between all three studies ($P < 0.001$ using one-way ANOVA), and was significantly different between each pair of studies ($P < 0.001$ for all pairwise comparisons using an unpaired t-test), although the average estimated hourly rates should have been approximately equal. Figure 8 illustrates the distribution of hourly rates across the studies. As can be seen, the hourly rates vary considerably, even though the data collection platform is the same, and the compensation levels and sampling criteria are similar. This can be attributed to different tasks, with Study 2 being the most demanding for the participants.

4.5.1. Study 1

The results in Figure 9 indicate that the flat rate used in Study 1 resulted in a higher calculated average hourly rate for all respondents ($n = 123$, 100%) than the estimated hourly rate of £8.91. The average actual hourly

Table 10. Summary of the included studies.

Study	N	Flat fee (£)	Minimum hourly wage (£)[a]	Actual hourly rate (£)	SD hourly rate (£)	CV	P
1	123	2.24	8.91	27.4	10.2	37.4%	<0.001
2	143	2	8.72	9.85	4.83	49%	0.005
3	97	3.5	8.72	13.7	5.89	43%	<0.001

P-values are based on comparing the minimum wage and the actual hourly rate. Statistical analysis was performed using one-sample t-test. The analysis was also repeated using median hourly rate, IQR, and one-sample Wilcoxon Signed-Rank test, which gave the same statistically significant differences.
[a]The UK minimum wage for people aged 23 and over was increased between the studies.

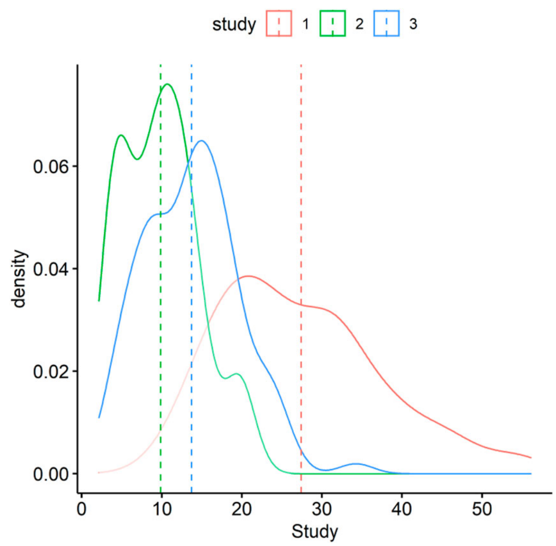

Figure 8. Distribution of hourly rates across studies. Lines represent the average calculated hourly rate.

rate (£27.4 ± £10.2) was significantly higher than the minimum hourly wage ($P < 0.001$). The figure communicates that using a flat rate is a bad idea because due to the variability in actual completion times, many crowdworkers end up being either over- or under-compensated relative to minimum wage. The previous analysis showed the average completion time to be 5.65 ± 2.3 min, which was 9.35 min shorter than the estimated time.

4.5.2. Study 2

In Study 2, the flat rate used resulted in an hourly rate greater than the estimated minimum wage in 57.3%

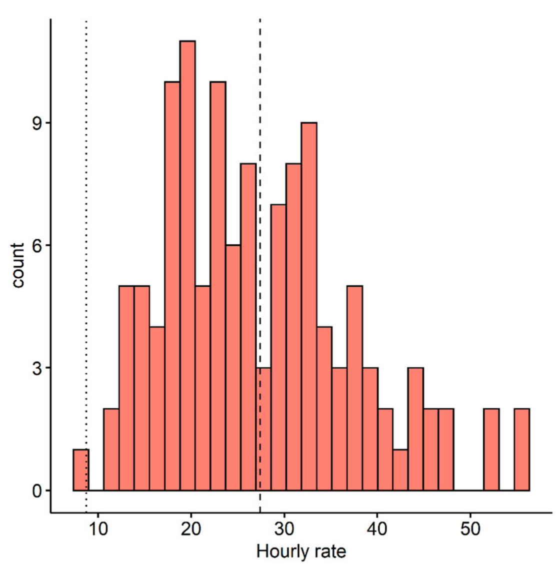

Figure 9. Distribution of the calculated hourly wage based on the flat rate and response time. The dashed line (right) represents the average actual hourly rate and the dotted line (left) represents the minimum wage.

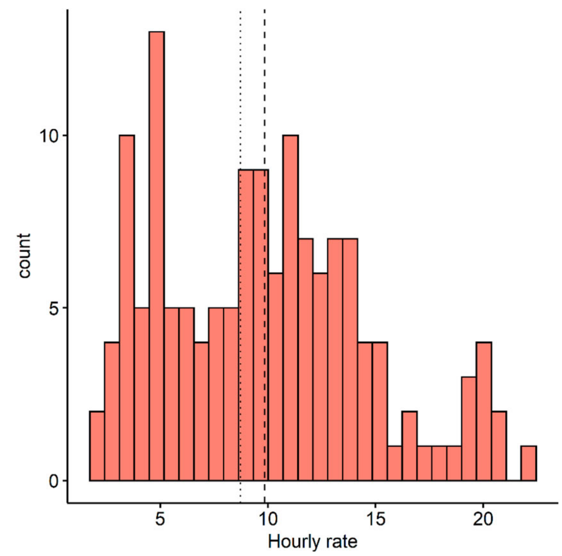

Figure 10. Distribution of the calculated hourly wage based on the flat rate and response time. Dashed line (right) represents the average actual hourly rate and the dotted line (left) represents the minimum wage.

($n = 82$) respondents and a lower hourly rate in the remaining 61 (42.7%), as shown in Figure 10. The figure shows that using a flat rate is sub-optimal because due to the variability in actual completion times, many crowdworkers end up being either over- or under-compensated relative to minimum wage. The average completion time was not significantly different from the estimated time at the 0.05 level ($P = 0.09$). Nonetheless, the average actual hourly rate (£9.85) was significantly higher than the estimated hourly rate of £8.72 ($P < 0.001$).

4.5.3. Study 3

In Study 3, the flat rate used resulted in an hourly rate greater than the estimated minimum wage in 74 (76.3%) respondents, and a lower hourly rate in the remaining 23 (23.7%), as shown in Figure 11. The figure indicates that using a flat rate is a sub-optimal idea, as due to the variability in actual completion times, many crowdworkers end up being either over- or under-compensated relative to minimum wage. The average completion time was significantly lower than the 25 min estimated completion time ($P < 0.001$). The average actual hourly rate (£13.7) was significantly higher than the estimated hourly rate of £8.72 ($P < 0.001$).

4.6. RQ6: Do researchers and workers win or lose by flat rate?

In Study 1, the total amount that should have been spent by the researchers if the actual distribution of time taken (based on the minimum wage) would have been £103

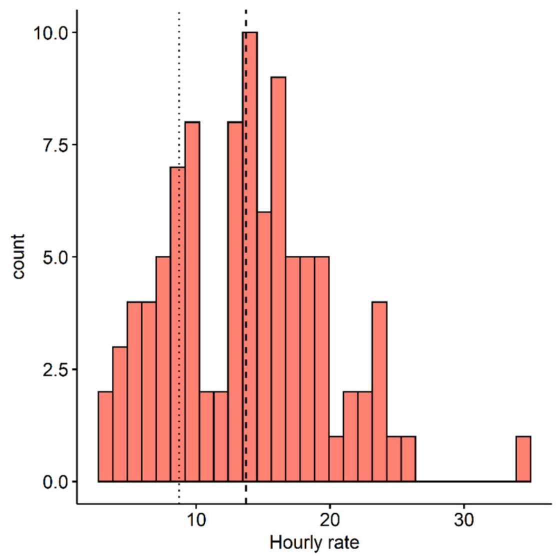

Figure 11. Distribution of the calculated hourly wage based on the flat rate and response time. The dashed line (right) represents the average actual hourly rate and the dotted line (left) represents the minimum wage.

(actual time × 8.91/60) while the estimated amount for the 123 participants should have been £276 (123 × £2.24). It can be concluded that *the researchers lost ~£173 due to the time variability*. In other words, the paid amount was 168% higher than it would have been based on actual completion times (given the same target hourly rate). So, for Study 1, the researchers lost in monetary terms. Symmetrically, the workers obtained a cumulative gain of the same amount (£173).

In Study 2, the total amount that should have been spent by the researcher if the actual distribution of time was taken into consideration (based on the minimum wage) would have been £342 (actual time × 8.72/60) while the estimated amount for the 143 participants should have been £286 (143 × £2). It can be concluded that *the researchers gained ~£50 due to the time variability*. In other words, the paid amount was 16% lower than it would have been, based on actual completion times and given the same target hourly rate. So, for Study 2, the researchers won in monetary terms. Symmetrically, the workers obtained a cumulative loss of the same amount (£50).

In Study 3, the total amount that should have been spent by the researchers if the actual distribution of time taken (based on the minimum wage) would have been £273 (actual time × 8.72/60), while the estimated amount for the 97 participants should have been £339 (97 × £3.5). It can be concluded that *the researchers lost ~£66 due to the time variability*. In other words, the paid amount was 24% higher than it would have been based on actual completion times (given the same target hourly rate). So, for Study 3, the researchers lost in monetary terms. Symmetrically, the workers obtained a cumulative gain of the same amount (£66).

Hence, in two studies, the researchers 'lost' and the crowdworkers 'gained' from the decided parameters of the estimated time of completion and fixed compensation. This was mainly due to the researchers overestimating (based on their pilot studies) how long it would actually take crowdworkers to complete the survey. In contrast, Study 2 turned out to be more time-demanding than expected based on its pilot study. It is relevant to point out that by adjusting the parameters, e.g. by mitigating the reward or time estimate, the outcome could have been different, and the crowdworkers would have 'lost' relative to the optimal situation where all crowdworkers are paid proportional to their actual time spent.

5. Discussion

5.1. Main implications

Our findings contribute to the nascent work of considering ethics in crowdwork (Irani and Six Silberman 2013; Irani and Six Silberman 2016; Whiting, Hugh, and Bernstein 2019). The findings show that even though it may appear tempting for researchers to set a flat rate for all crowdworkers, this is not the optimal strategy when the actual time taken varies by a considerable margin. Our results indicate that a non-trivial number of participants both exceed ('slow-runners') and go under the expected time ('speed-runners'), which reduces the usefulness of the flat rate approach and challenges researchers and platforms to develop better tools for fair compensation in crowdsourcing.

Our findings also underline the difficulty of estimating survey completion times when using online crowdsourcing platforms. In all studies, the assigned reward resulted in an actual hourly rate that was higher than the estimated hourly rate for more than 4 out of 5 participants. This is naturally a good outcome for the crowdworkers, but perhaps mitigates the cost-efficiency of data collection for the researchers. Previous research has found that targeting a specific wage is extremely difficult, and that researchers tend to underpay due to underestimating how long a task takes (Whiting, Hugh, and Bernstein 2019). While in our case the results also showed more overpayment relative to minimum wage, the conclusion of time-estimation being difficult remains the same.

Furthermore, the fact that the average hourly rate was significantly different between all three studies is both peculiar and interesting, and we explain it such that the inherent difficulty of the studies varied. Specifically, Study 2 had the lowest actual rate, indicating the highest difficulty. In contrast, Study 1 had the highest actual rate, indicating the lowest difficulty. In effect, the participants in Study 1 were paid a 178.2% higher hourly rate than those in Study 2 (although both groups were paid more than estimated rates). This difference underlines the importance of determining the actual compensations based on actual completion times, as estimating task difficulty is not a straightforward task (Cheng, Teevan, and Bernstein 2015).

Note that both under- and overcompensation can be considered to be unfair, but for different reasons. Undercompensation is unfair because it does not provide the target hourly wage for a worker. Overcompensation is also unfair because (a) it compensates Worker A more for their time than Worker B (even double or triple payment can be possible), and (b) it can reduce the total number of workers that are getting hired. This is because research budget B is divided between the number of participants as a function of the reward c, so that the N number of participants that can be recruited in a given budget is $N = B/c$. Hence, the larger the reward, the fewer people can be recruited. Consequently, overcompensation 'eats' the possibility of more people being able to earn.

Why is there time variability? We presume this is for the simple reason that the capabilities of the workers vary, and some are better at processing information quickly than others. Therefore, they can complete the same task significantly faster than those with average or relatively slow processing times, while the quality of the responses among these groups remains the same or at least at an acceptable level. Based on this premise, a fair system should not discriminate against slow workers – i.e. we should not penalise a person for being slow, given that they are doing their best in the work. However, the more there are ethical speed-runners (i.e. fast participants with good quality), the more likely it is that the ethical slowrunner *is* being penalised, when using averages to set the compensation level.

If the premise of time variability as a consequence of the innate abilities of workers holds, then it cannot be logically solved with any approach that relies on a constant, fixed compensation level for all. Instead, the realised compensation level needs to vary by participant, and be based on the empirical variability of their skills and abilities. This, we argue, is currently not realised in the majority of works conducted using crowdsourcing. Hence, it is imperative that researchers realise the importance of time variability and take the 'long path' of building time-aware compensation mechanisms, although we are the first to admit that this is challenging, given that it is exceedingly easy to rely on the standard features provided by the crowdsourcing platforms due to their convenience and speed.

5.2. Implications for researchers

As human computation systems become more common, there is a desire among researchers to offer fair compensation (Alonso, Marshall, and Najork 2013; Barbosa and Chen 2019; Irani and Six Silberman 2013; Whiting, Hugh, and Bernstein 2019; Silberman et al. 2018) to crowdworkers feeding data to these systems, as well as to those providing valuable inputs for survey research. Unfortunately, current platforms lack easy tools for assigning fair compensation, and there are no ready mechanisms to avoid under- or overcompensation. In the following, we suggest three basic strategies for researchers to provide fair compensation in crowdsourcing based on the time implications raised.

Strategy 1: Compensation by Normal Distribution. Intuitively, the idea here is to, in a way, overestimate the time taken to complete the study. For example, first carry out a pilot study with N participants (an exploratory fraction of the target sample size). This helps to obtain a mean value (μ), i.e. the average time it took the pilot participants to complete the study, as well as their SD. Then, assuming an hourly compensation goal (CG), the reward (R) can be computed as follows: $R = CG \times [(\mu + 3 \times SD)/60]$. For example, assume we want every participant to be paid at least 10 USD per hour. Then, based on a pilot run we observe that on average participants take 15 min to complete the task, with the SD being 5 min. We then simply add another 15 min ($3 \times 5 = 15$) to our expected completion time, yielding a 'safe' estimate of 30 min, according to which we set the task completion reward (which in this case would be 5 USD). This strategy can be applied when it is not possible to alter the paid rewards after task completion. Theoretically, this strategy increases the chances that a crowdworker may be underpaid, as according to the assumption of normal distribution, 99.7% of the duration values would be located within three SDs of the mean. Naturally, this comes at the cost of overcompensating most of the workers that lie nearer to the average completion time. However, this strategy can be applied when aiming to minimise the chances of unfair compensation across all workers. From a cost–benefit perspective, one can make the argument that this approach is still beneficial for the researcher when comparing the cost of paying slightly more than minimum wage with the researcher career benefits of a published peer-reviewed publication or similar professional outcome.

Strategy 2: Multi-Objective Fairness. Theoretically, it is possible to build a custom solution that considers time variability, as well as other relevant factors such as output quality and the crowdworker's physical location. Then, the general formula of setting the reward for Participant i is as follows: $R_i = quality_i \times time_i \times location_i$. In other words, given that a participant passes the quality check, the reward is set based on the time they took to complete the task and their local minimum wage (or a variant of it, e.g. a heuristically inferred rule of *X% above minimum wage*). This requires that the researcher tracks the time taken and obtains the geographic location of the participant. In theory, this strategy has the negative effect of incentivising slow work (Whiting, Hugh, and Bernstein 2019) if the worker knows that they will be compensated more if they work slowly. Hence, perhaps not revealing that the time taken is a factor for compensation might be more beneficial for obtaining durations that match perfectly with the real capabilities of the workers. Additionally, the measurement of time could incorporate the invisible work done by the workers (Toxtli, Suri, and Savage 2021), and not only the task completion time.

Strategy 3: Post-Hoc Bonuses. This strategy involves observing the participants' actual completion times, and allocating a bonus to those falling behind the determined minimum wage or other objective level of compensation. Theoretically, as this approach is data-driven, it ensures that all participants reach the minimum required compensation. The downside of this approach is that it is labourious, requires manual calculations, and it also requires that the platform enables sending bonuses after task completion. On the other hand, the calculation is not difficult: $g_i = hourly\ minimum\ wage - 1/actual\ hourly\ time\ taken_i \times reward$ (see example[1]), giving the wage gap g of participant i. The gaps can then be paid using the platform's bonus feature (which is available on some current platforms, Wang, Yang, and Wang 2019). However, it appears that the calculations would need to be done only on a subset of actual participants, as most crowdworkers finish the task under time.

When addressing fairness, there should also be some consideration given to the workers' cost of living. Namely, if applying the factor of geographic location, we can pose the following question: *Should a person from lower-income country be paid less just because they come from a lower-income country?* There can be two answers: yes and no. Yes, because if the worker is *not* paid a sum that is relative to their local living standard, then this may result in unfair treatment of the workers. For example, presume Participant A has a local minimum wage of \$5, and Participant B has a local minimum wage of \$10. Now, if both take the same time to complete the study successfully and get the same compensation, this compensation y will always be twice as favourable to Participant A ($(y/5)/(y/10) = 2$) because they live in a country where the same sum can provide more in general. Note that this location premise is also behind remote work compensation schemes – for example, Google setting their employees' salaries based on their physical location.[2] On the other hand, there is a contrary argument, that in online environments, people *should* be paid the same for the same job, regardless of their country of origin, as the virtual work market is global. According to this premise and rooted in the principle of equal opportunity (Busarovs 2013), compensation is fair when matching outputs are appraised equally. While this can result in people from lower-income countries gaining a relative advantage over those from higher-income countries, this is accepted based on the quality of work (i.e. the fact that the intrinsic value for the researcher is the same).

Finally, our findings underline the difficulty of estimating task completion times, which is a problem identified in previous work (Cheng, Teevan, and Bernstein 2015; Saito et al. 2019) and which is confirmed by our findings. Possible reasons for this include inaccurate estimation services being provided by the crowd working platform, and/or the participants taking part in the pilot testing of the tasks not being representative of the crowdworker population who actually take on the task. However, based on the three studies we analyzed there does seem to be a visible trend, where the more complex the task (which for this research was measured by the number of questions on the survey), the more convergent were the actual task time and the estimated task time.

5.3. Implications for crowdsourcing platforms

In the wake of worker-centered design (Fox et al. 2020) that aims to support (online) labour, we offer three design propositions (DPs) that could increase compensation fairness:

- **DP01: Conducting post-hoc analysis.** Platforms could show researchers information about participants that fell under the minimum wage based on the actual completion times, and give researchers the option to easily increase their compensation to the minimum wage (or another desired level). Currently, we are not aware of any platforms that have this feature.
- **DP02: Mitigating information asymmetry.** The current platforms offer researchers little information

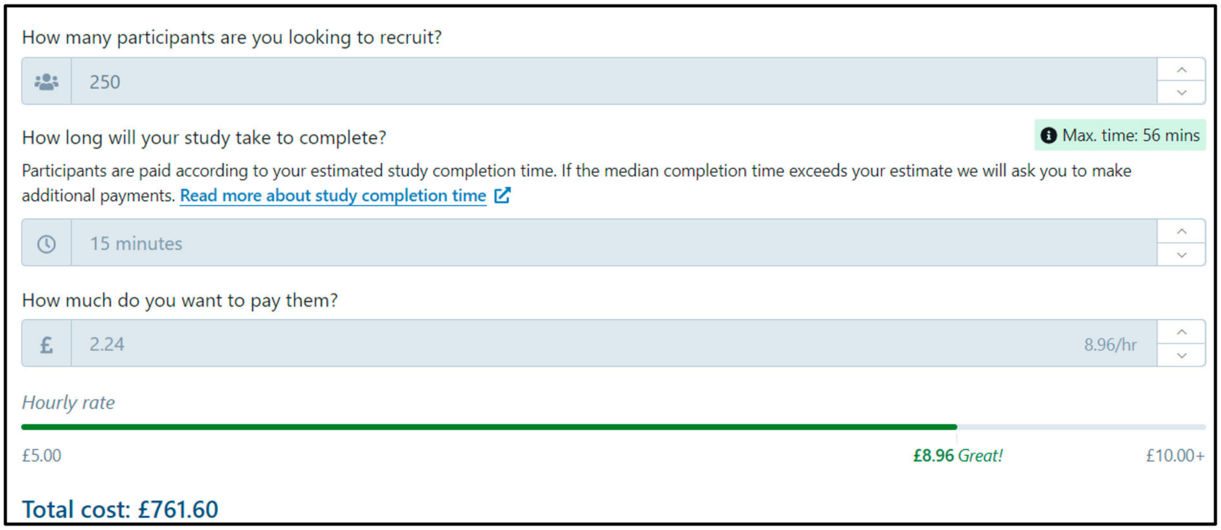

Figure 12. Example of platform functionalities that support calculating fair hourly rates (Prolific, Study 1). By changing either the estimated completion time (the field in the middle) or the compensation level (the field at the bottom), the system calculates the estimated hourly rate (the green bar at the bottom), and gives a visual and textual indication of how fair it is (in this case, '£8.96 Great!').

about fair compensation. Among the missing information are (a) *suggestions* for target minimum wages, and (b) *notifications* of under- or over-payment in real-time. Platforms could provide these features as an added form of system transparency (Garfinkel et al. 2017). A relevant point here is that novice and expert crowdworkers' evaluations of task feasibility differ (Hanrahan et al. 2021), implying that these two worker groups might require specialised solutions.

- **DP03: Applying an auction mechanism.** The payment level of the researcher could be adjusted in real-time, based on actual completion times (similar to the auctions taking place in online advertising, Jansen and Mullen 2008). For example, if the task is taking longer than expected, then the flat rate increases up to a researcher-set maximum (i.e. a cap price).

Obviously, more research and development is needed to design fair compensation functionalities for use in crowdsourcing platforms (Gaikwad et al. 2015; Irani and Six Silberman 2013; Salehi et al. 2015). Current platforms, such as MTurk, Appen, and Prolific do not adequately address this issue. Both MTurk and Appen lack functionalities for setting a fair level of compensation, essentially 'pushing the problem' onto the researchers. Prolific has features that facilitate calculating an hourly wage (see Figure 12), and if the researcher knows the local minimum wage and focuses on a geographic region, these functionalities help approximate fair wages. However, as we have shown in this study, a fixed hourly wage suffers from the problem of averaging, and therefore does not provide a reliable proxy for fair compensation at the individual level.

Therefore, the question is: *How should a crowdsourcing platform be designed for fair compensation?* Or more succinctly, *how can a fair platform be designed?* Naturally, these question is far from being trivial, as evidenced by the large amount of work carried out in this field. The approach to fairness is ultimately a choice. As an example, consider two cashiers, one fast and one slow. Should the slow cashier be paid less than the fast for the same hours? In most companies, people are paid the same amount for the same job regardless of speed. Then, there are other professions where performance-based compensation is more accepted, for example, sales. Typically, corporate salespeople have some base salary and then performance-based bonuses. The question is – is crowdsourcing more like working on corporate sales or in a supermarket? This question might not be for the researchers to decide, but the answer might require a policymaker intervention.

The challenge is partly due to the requirement of catering to the needs of all stakeholders in a fair manner. The platforms may perceive their predominant incentive to be directed towards researchers, as the researchers are the source of funds, and therefore, possibly seen as less expendable than a typical 'small' crowdworker, of whome there are potentially millions. On the other hand, platforms may be aware of the need for quality workers, as low levels of quality would rapidly result in researchers abandoning the platform. Hence, recognising the nature of crowdsourcing platforms as two-sided markets (Rochet and Tirole 2003) seems like a necessary starting point for designing fairness mechanisms. To provide fairness for researchers, the platforms should prevent scenarios where the researcher is overcharged – i.e. having to pay more than could have been paid for the same quality of work. To provide fairness for workers, compensation levels should not fall under the expected fair price, but as mentioned, this is a tricky proposition because the motivations for participating in crowdwork can vary from simply passing time to supporting a family (Brewer, Morris, and Piper 2016; Kaplan et al. 2018; Kaufmann, Schulze, and Veit 2011), thus resulting in very different expectations.

Perhaps somewhat unexpectedly, overpayment can also decrease the overall fairness in the system. This is because the more a researcher overpays one worker, the less budget they have for other workers. Thus, the scenario arises where if one could have hired three people and satisfied their minimum living costs (for which the minimum wage is a proxy), by overpaying, one might only have the capacity to hire a single worker with the same money. Certainly, it is good for the person who receives the overpayment, but it is bad for others that are left without any chance for earning as a consequence.

Navigating the different expectations and needs of stakeholders can be demanding. Researchers may prefer an estimate from the platform of how much their overall data collection would cost. But obviously, if the cost is time-variable and the expected time cannot be computed reliably, this can cause the researcher to exceed their budget, thus lowering their satisfaction with the platform and possibly resulting in incomplete datasets. Moreover, if the compensation is tied to the time of completion, then there is an incentive for the crowdworkers to complete the study slower than normally. So, while the general goal should be to factor in the natural speed differences among workers, the question of *how* to do so remains a concrete problem to tackle.

5.4. Policy implications

It can also be argued that neither researchers nor platforms should determine the rules of crowdwork. Especially, platforms are private enterprises whose primary allegiance is to the maximisation of profits for shareholders. Furthermore, leaving the question of fairness down to the researcher involves the complication that each researcher might have a very different definition of fairness. All of these definitions could be equally defendable. For example, time-based fairness might add inequity between workers where one crowdworker is completing the task faster because he is working harder, and another slower because he is distracted. Therefore, on this ground, it would be justifiable to subscribe to performance-based fairness instead of time-based fairness.

The approach to fairness is ultimately a choice, and in most companies, people are paid the same amount regardless of the speed at which they work. Yet, there are other professions where performance-based compensation is more accepted, for example, sales. Typically, corporate salespeople have a base salary and then recieve performance-based bonuses. This raises the question of whether crowdsourcing more like working in corporate sales, or working in a supermarket? Because researchers' answers to this question might be opinionated, or researchers might not be the right people to answer it in the first place, fairness might be very difficult to determine within the system. Hence, perhaps fairness needs to be determined by policymakers.

To some extent, policy is already made. For example, US law requires that implementers do not provide specific payment constraints, or else the categorisation of the platform workers may be obliged to be changed. In a similar vein, universities' Institutional Review Boards, the National Science Foundation, or publishers could determine a fair compensation policy for crowdwork, be it minimum wage or some other guideline for researchers.

5.5. Limitations and future work

As mentioned in the discussion section, fairness may be defined in several ways: for example, either ignoring or factoring in criteria for fairness such as the time taken to complete the task, or the participant's cost of living. In this study and in line with previous studies (Whiting, Hugh, and Bernstein 2019), we did not adjust the hourly rates by the participants' location, therefore, adhering to the first definition of fairness. However, for achieving comprehensiveness, it would be possible to compute the impact of a worker's country on fair compensation. This can be done simply through the formula $c_i = actual\ time\ taken_i * hourly\ local\ minimum\ wage_i$, indicating the compensation c of participant i. After computing these values for a set of participants, we can further compare how much the paid compensation differs from locally adjusted hourly wages.

Also, we focused mostly on survey research which involved studies that asked multiple questions of a different nature from the participants (i.e. not a repetitive task). This study type is different from other crowdsourced study types such as machine learning annotation (Salminen et al. 2018b; Weber and Mejova

2016) or HITs (Alonso 2015), which, for example, might ask participants to identify if a picture contains a certain object. HITs in general are considered less demanding than subjective tasks such as survey completion that requires more elaborate thinking. Therefore, our findings may not apply to all types of crowdsourced tasks, but they do apply to using crowdsourcing for surveys such as those commonly carried out in user studies and social science research (e.g. psychology, sociology). Therefore, it would be interesting to see the compensation expectations from crowdworkers in these types of HIT tasks (see Kutlu et al. 2020).

In our study, we used minimum wage as a proxy for fair compensation. However, it is noteworthy to mention that minimum wage may not necessarily correlate with the crowdworkers' perceptions of fair compensation, nor correlate with notions of fair task compensation. For example, Whiting, Hugh, and Bernstein (2019) found that workers considered fair compensation to be double the US federal minimum wage at the time. This puts pressure on researchers to not just meet the minimum wage, but also to find ways of exceeding it in order to satisfy crowdworkers' sense of fair compensation. Furthermore, this challenges the idea of using the minimum wage as a reliable baseline for fair compensation. If the data provided by the workers is critical for the conduct of the research, and the research is funded with sizeable research grants, then it is fair to ask *whether the minimum wage is too low?* Moreover, an aspect that we did not consider but that matters for workers is their 'invisible work', namely locating tasks, communicating with researchers, and managing payments (Toxtli, Suri, and Savage 2021). These tasks effectively decrease the hourly wage of the workers, and should be considered in future analysis.

It appears that optimising one notion of fairness can undermine another, especially in terms of skills versus time. For example, slower performance might be due to genuine (e.g. impairment) or ingenuine (e.g. distraction) reasons (Zhang et al. 2022). Hence, it would be valuable to examine this tradeoff more systematically, for example, by evaluating how design factors in payment mechanisms contribute to different notions of fairness, and how these notions are valued by relevant stakeholders (e.g. workers, researchers, and platforms). In a similar vein, future work should examine the relationship between workers' speed and their level of expertise (Hanrahan et al. 2021).

Finally, our results are based on three studies conducted on the same crowdsourcing platform. Therefore, it is possible that they reflect the response patterns of participants on that platform, and that those patterns might deviate from patterns observed on other platforms. We chose Prolific because of its out-of-the-box feature for recording the time taken to complete a study. However, ideally, future work can compare our results with those obtained from other platforms. While the findings may change in terms of speed-runner/slow-runner ratios, we believe the main finding is that the time taken to complete surveys varies greatly among crowdsourcing participants, and that this has direct and important implications for fair compensation.

6. Conclusion

Our findings highlight the complexities of determining fair compensation for crowdwork at aggregate or individual levels, and for either the crowdworker or the researcher. Using one standard rate of compensation suffers from the problem of averaging, and therefore does not provide a reliable proxy for fair compensation. We suggest that researchers set a fair level of compensation that addresses factors such as the quality of work, time variability based on workers' natural abilities, and also geographic location (i.e. local standard of living).

Notes

1. For example, hourly minimum wage = 8.00; time taken = 0.5 h; reward = 2.00 → $g = 8 - (1/0.5 * 2) = 4\$$.
2. https://www.hrkatha.com/news/compensation-benefits/googlers-can-use-work-location-tool-to-calculate-pay-for-remote-work/.

Acknowledgements

Open Access funding provided by the Qatar National Library.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Finnish Foundation for Economic Education [grant number 210003].

ORCID

Ahmed Mohamed Sayed Kamel http://orcid.org/0000-0002-3791-5998
Mekhail Mustak http://orcid.org/0000-0002-2111-2939

References

Achrol, Ravi S., and Gregory T. Gundlach. 1999. "Legal and Social Safeguards Against Opportunism in Exchange." *Journal of Retailing* 75 (1): 107–124.

Alonso, Omar. 2015. "Practical Lessons for Gathering Quality Labels at Scale." In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (SIGIR '15), ACM, New York, NY, USA, 1089–1092. doi:10.1145/2766462.2776778.

Alonso, Omar, Catherine C. Marshall, and Marc A. Najork. 2013. "A Human-Centered Framework for Ensuring Reliability on Crowdsourced Labeling Tasks." In *First AAAI Conference on Human Computation and Crowdsourcing*. Accessed 27 March 2018. https://www.aaai.org/ocs/index.php/HCOMP/HCOMP13/paper/view/7487.

Alonso, Omar, Daniel E. Rose, and Benjamin Stewart. 2008. "Crowdsourcing for Relevance Evaluation." In *ACM SigIR Forum*, edited by Björn Hartman and Eric Horvitz, 9–15. New York: ACM.

Barbosa, Natã M., and Monchu Chen. 2019. "Rehumanized Crowdsourcing: A Labeling Framework Addressing Bias and Ethics in Machine Learning." In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12.

Brewer, Robin, Meredith Ringel Morris, and Anne Marie Piper. 2016. "Why Would Anybody Do This? Understanding Older Adults' Motivations and Challenges in Crowd Work." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2246–2257.

Busarovs, Aleksejs. 2013. "Ethical Aspects of Crowdsourcing, or is it a Modern Form of Exploitation." *International Journal of Economics & Business Administration.* 1 (1): 3–14.

Cao, Alex, Keshav K. Chintamani, Abhilash K. Pandya, and R. Darin Ellis. 2009. "NASA TLX: Software for Assessing Subjective Mental Workload." *Behavior Research Methods* 41 (1): 113–117.

Cheng, Justin, Jaime Teevan, and Michael S. Bernstein. 2015. "Measuring Crowdsourcing Effort With Error-Time Curves." In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1365–1374. Accessed 4 July 2021. doi:10.1145/2702123.2702145.

Davidson, Thomas, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. "Automated Hate Speech Detection and the Problem of Offensive Language." In *Proceedings of Eleventh International AAAI Conference on Web and Social Media*, Montreal, Canada, 512–515.

Difallah, Djellel, Elena Filatova, and Panos Ipeirotis. 2018. "Demographics and Dynamics of Mechanical Turk Workers." In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining* (WSDM '18), Association for Computing Machinery, New York, NY, USA, 135–143. doi:10.1145/3159652.3159661.

Erickson, Ingrid, Mohammad H. Jarrahi, Leslie Thomson, and Steve Sawyer. 2014. "More Than Nomads: Mobility, Knowledge Work, and Infrastructure." In *Proceedings of the European Group for Organizational Studies Colloquium*.

Fox, Sarah E., Vera Khovanskaya, Clara Crivellaro, Niloufar Salehi, Lynn Dombrowski, Chinmay Kulkarni, Lilly Irani, and Jodi Forlizzi. 2020. "Worker-Centered Design: Expanding HCI Methods for Supporting Labor." In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–8.

Gaikwad, Snehal, Durim Morina, Rohit Nistala, Megha Agarwal, Alison Cossette, Radhika Bhanu, Saiph Savage, Vishwajeet Narwal, Karan Rajpal, and Jeff Regino. 2015. "Daemo: A Self-Governed Crowdsourcing Marketplace." In *Adjunct Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*, 101–102.

Garfinkel, Simson, Jeanna Matthews, Stuart S. Shapiro, and Jonathan M. Smith. 2017. "Toward Algorithmic Transparency and Accountability." *Communications of the ACM* 60 (9): 5. doi:10.1145/3125780.

Gosling, Samuel D., Peter J. Rentfrow, and William B. Swann. 2003. "A Very Brief Measure of the Big-Five Personality Domains." *Journal of Research in Personality* 37 (6): 504–528.

Gray, Mary L., and Siddharth Suri. 2019. *Ghost Work: How to Stop Silicon Valley from Building a New Global Underclass*. Boston and New York: Eamon Dolan Books.

Hanrahan, Benjamin V., Anita Chen, JiaHua Ma, Ning F. Ma, Anna Squicciarini, and Saiph Savage. 2021. "The Expertise Involved in Deciding Which HITs are Worth Doing on Amazon Mechanical Turk." *Proceedings of the ACM on Human-Computer Interaction* 5: 1–23.

Hara, Kotaro, Abigail Adams, Kristy Milland, Saiph Savage, Chris Callison-Burch, and Jeffrey P. Bigham. 2018. "A Data-Driven Analysis of Workers' Earnings on Amazon Mechanical Turk." In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–14.

Hong, Yili, and Paul Pavlou. 2012. "An Empirical Investigation on Provider Pricing in Online Crowdsourcing Markets for IT Services." In *Proceedings of the Thirty Third International Conference on Information Systems*, Orlando, Florida, USA.

Irani, Lilly C., and M. Six Silberman. 2013. "Turkopticon: Interrupting Worker Invisibility in Amazon Mechanical Turk." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 611–620. Accessed 4 July 2021. doi:10.1145/2470654.2470742.

Irani, Lilly C., and M. Six Silberman. 2016. "Stories We Tell About Labor: Turkopticon and the Trouble with Design." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, Association for Computing Machinery, New York, NY, USA, 4573–4586. Accessed 4 July 2021. doi:10.1145/2858036.2858592.

Jäätmaa, Jaakko. 2010. "Financial Aspects of Cloud Computing Business Models." Master's thesis. Aalto University, Helsinki, Finland.

Jansen, Bernard J., and Tracy Mullen. 2008. "Sponsored Search: An Overview of the Concept, History, and Technology." *International Journal of Electronic Business* 6 (2): 114–131.

Kaplan, Toni, Susumu Saito, Kotaro Hara, and Jeffrey P. Bigham. 2018. "Striving to Earn More: A Survey of Work Strategies and Tool Use Among Crowd Workers." In

Sixth AAAI Conference on Human Computation and Crowdsourcing.

Kaufmann, Nicolas, Thimo Schulze, and Daniel Veit. 2011. "More Than Fun and Money: Worker Motivation in Crowdsourcing-A Study on Mechanical Turk." In *AMCIS*, Detroit, USA, 1–11.

Kittur, Aniket, Ed H. Chi, and Bongwon Suh. 2008. "Crowdsourcing User Studies With Mechanical Turk." In *Proceedings of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems* (CHI '08), ACM, New York, NY, USA, 453–456. doi:10.1145/1357054.1357127.

Kothe, Emily, and Mathew Ling. 2019. "Retention of Participants Recruited to a One-Year Longitudinal Study Via Prolific." doi:10.31234/osf.io/5yv2u.

Kutlu, Mucahid, Tyler McDonnell, Tamer Elsayed, and Matthew Lease. 2020. "Annotator Rationales for Labeling Tasks in Crowdsourcing." *Journal of Artificial Intelligence Research* 69: 143–189. doi:10.1613/jair.1.12012.

Mahtani, Kamal, Elizabeth A. Spencer, Jon Brassey, and Carl Heneghan. 2018. "Catalogue of Bias: Observer Bias." *BMJ Evidence-Based Medicine* 23 (1): 23–24.

Mankar, Akash, Riddhi J. Shah, and Matthew Lease. 2017. "Design Activism for Minimum Wage Crowd Work." *ArXiv170610097 Cs*. Accessed 4 July 2021. http://arxiv.org/abs/1706.10097.

Mao, Andrew, Ece Kamar, Yiling Chen, Eric Horvitz, Megan E. Schwamb, Chris J. Lintott, and Arfon M. Smith. 2013a. "Volunteering Versus Work for Pay: Incentives and Tradeoffs in Crowdsourcing." In *First AAAI Conference on Human Computation and Crowdsourcing*.

Mao, Ke, Ye Yang, Mingshu Li, and Mark Harman. 2013b. "Pricing Crowdsourcing-Based Software Development Tasks." In 2013 *35th International Conference on Software Engineering (ICSE)*, IEEE, 1205–1208.

McInnis, Brian, Dan Cosley, Chaebong Nam, and Gilly Leshed. 2016. "Taking a HIT: Designing Around Rejection, Mistrust, Risk, and Workers' Experiences in Amazon Mechanical Turk." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2271–2282.

Morrell, Daniel L., Jeffrey B. Griffin, and Millicent Nelson. 2018. "Crowdsourcing as a Career Trend: Are Online Workers the New Contingent Workforce." *Journal of Strategic Human Resource Management* 7 (3): 1.

Narula, Prayag, Philipp Gutheim, David Rolnitzky, Anand Kulkarni, and Bjoern Hartmann. 2011. "Mobileworks: A Mobile Crowdsourcing Platform for Workers at the Bottom of the Pyramid." In *Workshops at the Twenty-Fifth AAAI Conference on Artificial Intelligence*.

Palan, Stefan, and Christian Schitter. 2018. "Prolific. AC – A Subject Pool for Online Experiments." *Journal of Behavioral and Experimental Finance* 17 (2018): 22–27.

Peer, Eyal, Laura Brandimarte, Sonam Samat, and Alessandro Acquisti. 2017. "Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research." *Journal of Experimental Social Psychology* 70 (2017): 153–163.

Rochet, Jean-Charles, and Jean Tirole. 2003. "Platform Competition in Two-Sided Markets." *Journal of the European Economic Association* 1 (4): 990–1029.

Saito, Susumu, Chun-Wei Chiang, Saiph Savage, Teppei Nakano, Tetsunori Kobayashi, and Jeffrey P. Bigham. 2019. "TurkScanner: Predicting the Hourly Wage of Microtasks." In *The World Wide Web Conference* (WWW '19), Association for Computing Machinery, New York, NY, USA, 3187–3193. doi:10.1145/3308558.3313716.

Salehi, Niloufar, Lilly C. Irani, Michael S. Bernstein, Ali Alkhatib, Eva Ogbe, and Kristy Milland. 2015. "We are Dynamo: Overcoming Stalling and Friction in Collective Action for Crowd Workers." In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, 1621–1630.

Salminen, Joni, Hind Almerekhi, Milica Milenković, Soon-gyo Jung, Jisun An, Haewoon Kwak, and Bernard J. Jansen. 2018a. "Anatomy of Online Hate: Developing a Taxonomy and Machine Learning Models for Identifying and Classifying Hate in Online News Media." In *Proceedings of The International AAAI Conference on Web and Social Media (ICWSM 2018)*, San Francisco, California, USA. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM18/paper/view/17885.

Salminen, Joni, Lene Nielsen, Soon-Gyo Jung, and Bernard J. Jansen. 2021. "Towards a Measurement Scale of Organizational Readiness for Personas." In *Extended Abstracts of CHI Conference on Human Factors in Computing Systems (CHI2021)*, ACM, Virtual Conference.

Salminen, Joni, Joao M. Santos, Soon-gyo Jung, Motahhare Eslami, and Bernard J. Jansen. 2019. "Personal Transparency: Analyzing the Impact of Explanations on Perceptions of Data-Driven Personas." *International Journal of Human–Computer Interaction* 0 (0): 1–13. doi:10.1080/10447318.2019.1688946.

Salminen, Joni, Joao M. Santos, Haewoon Kwak, Jisun An, Soon-gyo Jung, and Bernard J. Jansen. 2020. "Persona Perception Scale: Development and Exploratory Validation of an Instrument for Evaluating Individuals' Perceptions of Personas." *International Journal of Human-Computer Studies* 141: 102437. doi:10.1016/j.ijhcs.2020.102437.

Salminen, Joni, Fabio Veronesi, Hind Almerekhi, Soon-gyo Jung, and Bernard J Jansen. 2018b. "Online Hate Interpretation Varies by Country, But More by Individual: A Statistical Analysis Using Crowdsourced Ratings." In *Proceedings of The Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS – 2018)*, IEEE, Valencia, Spain.

Savage, Saiph, Chun Wei Chiang, Susumu Saito, Carlos Toxtli, and Jeffrey Bigham. 2020. "Becoming the Super Turker: Increasing Wages Via a Strategy from High Earning Workers." In *Proceedings of The Web Conference* 2020, 1241–1252.

Seidman, Gay W. 2007. *Beyond the Boycott: Labor Rights, Human Rights, and Transnational Activism*. New York: Russell Sage Foundation.

Sheng, Victor S., Foster Provost, and Panagiotis G. Ipeirotis. 2008. "Get Another Label? Improving Data Quality and Data Mining Using Multiple, Noisy Labelers." In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery And Data Mining*, ACM, 614–622.

Silberman, M. S., B. Zaldivar, R. Tomlinson, J. LaPlante, L. Ross, and A. Irani. 2018. "Responsible Research with Crowds: Pay Crowdworkers at Least Minimum Wage." *Communications of the ACM 61*. Accessed 4 July 2021. https://cacm.acm.org/magazines/2018/3/225476-responsible-research-with-crowds/fulltext.

Singer, Yaron, and Manas Mittal. 2013. “Pricing Mechanisms for Crowdsourcing Markets.” In *Proceedings of the 22nd International Conference on World Wide Web*, ACM, 1157–1166.

Slivkins, Aleksandrs, and Jennifer Wortman Vaughan. 2014. “Online Decision Making in Crowdsourcing Markets: Theoretical Challenges.” *ACM SIGecom Exchanges.* 12 (2): 4–23.

Stigler, George J. 1946. “The Economics of Minimum Wage Legislation.” *The American Economic Review* 36 (3): 358–365.

Sutherland, Will, Mohammad Hossein Jarrahi, Michael Dunn, and Sarah Beth Nelson. 2020. “Work Precarity and Gig Literacies in Online Freelancing.” *Work, Employment and Society* 34 (3): 457–475.

Thabane, Lehana, Jinhui Ma, Rong Chu, Ji Cheng, Afisi Ismaila, Lorena P. Rios, Reid Robson, Marroon Thabane, Lora Giangregorio, and Charles H. Goldsmith. 2010. “A Tutorial on Pilot Studies: The What, Why and How.” *BMC Medical Research Methodology* 10 (1): 1–10.

Tong, Yongxin, Libin Wang, Zimu Zhou, Lei Chen, Bowen Du, and Jieping Ye. 2018. “Dynamic Pricing in Spatial Crowdsourcing: A Matching-Based Approach.” In *Proceedings of the 2018 International Conference on Management of Data*, 773–788.

Toxtli, Carlos, Siddharth Suri, and Saiph Savage. 2021. “Quantifying the Invisible Labor in Crowd Work.” *Proceedings of the ACM on Human-Computer Interaction* 5: 1–26.

Wang, Lili, Ye Yang, and Yong Wang. 2019. “Do Higher Incentives Lead to Better Performance? An Exploratory Study on Software Crowdsourcing.” In *2019 ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, IEEE, 1–11.

Weber, Ingmar, and Yelena Mejova. 2016. “Crowdsourcing Health Labels: Inferring Body Weight from Profile Pictures.” In *Proceedings of the 6th International Conference on Digital Health Conference*, 105–109.

Whiting, Mark E., Grant Hugh, and Michael S. Bernstein. 2019. “Fair Work: Crowd Work Minimum Wage With One Line of Code.” In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 197–206.

Zhang, Xuanhui, Weijia Zhang, Yuxiang Chris Zhao, and Qinghua Zhu. 2022. “Imbalanced Volunteer Engagement in Cultural Heritage Crowdsourcing: A Task-Related Exploration Based on Causal Inference.” *Information Processing & Management* 59 (5): 103027. doi:10.1016/j.ipm.2022.103027.