Prolific ID

## Consent

This task is part of a voluntary research study conducted by Carmel Baharav, Jingyan Wang, Lily Laredo, R Ravi, Nihar Shah, and Anita Woolley at Carnegie Mellon University and funded by the Block Center for Technology and Society at Carnegie Mellon University. We can't tell you the full purpose of the research in advance, but we'll debrief you at the end of the study. After the study is complete, you will have the opportunity to send us your comments or contact us to ask questions.

In this task, you will answer several questions. These questions involve looking at a sample of programming challenge scores taken from resumes and judging how strong/weak the candidates are compared to what you know about the rest of the applicant pool. This task will take approximately 5 minutes. You will receive $0.8 in base compensation, and a potential bonus of $0.2 based on your performance on the task within 3 days of completion. Participation in this study is limited to individuals age 18 and older.

There is a potential for breach of confidentiality, in that Prolific links your response to your Worker ID. To minimize this risk, we will immediately delete Worker IDs and will not store them on our computers. Otherwise, the study does not introduce any risk which is greater than those ordinarily encountered on Prolific and there may be no personal benefit from your participation in the study. We will remove personal identifiers from collected responses and will only store and share de-identified data. De-identified data may be made available to other researchers once our papers are published.

If you have any questions about this study, please contact Jingyan Wang now at jingyanw@cs.cmu.edu. If you have questions later, desire additional information, or wish to withdraw your participation, please contact Jingyan at the email above.

If you have questions pertaining to your rights as a research participant; or to report concerns to this study, you should contact the Office of Research integrity and Compliance at Carnegie Mellon University. Email: irb-review@andrew.cmu.edu. Phone: 412-268-1901 or 412-268-5460.

Your participation in this research is voluntary. You may discontinue participation at any time.

By clicking on the "I agree" button, I confirm that I am at least 18, that I have read and understand the above information, and that I wish to participate in the research and begin the task.

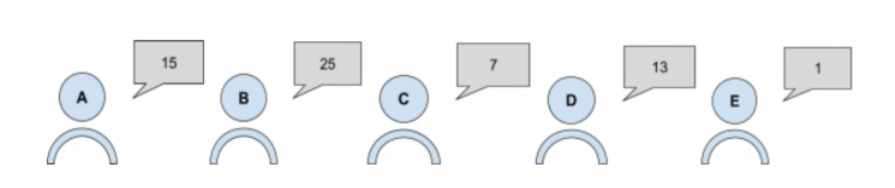
☐ I agree

## Introduction

You are a hiring manager at a firm. There are 1000 applicants applying for a job. All job applicants must complete a test as part of their application. You only see scores from a small random set of applicants. **Your task is to put each applicant in a category, based on how you think they compare to the others among the 1000 applicants**. There are five categories: 80-100% (best), 60-80%, 40-60% (average), 20-40%, 0-20% (worst). For example, being 60-80% means you believe this person is above 60% of the people, and below 20% of the people.

## Demo

Think about the following example. Applicants took a test, and received a score of 0-100, where higher is better. You see the following five applicants and their scores:

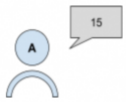
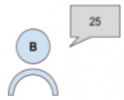
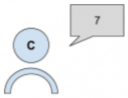
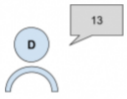
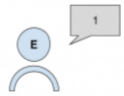



At first you may think 15 and 25 are low, given scores could be as high as 100, but then you see that compared to the rest of the scores, 15 and 25 are actually pretty high. So you put them in the following categories:


Then you learn that the test for the applicants was actually a pretty hard test, so your guesses were close to accurate. In reality, the scores of the 1000 applicants looked like this:

| Category | Scores |
|---|---|
| 0-20% (worst) | 0-8 |
| 20-40% | 9-12 |
| 40-60% (average) | 13-14 |
| 60-80% | 15-25 |
| 80-100% (best) | 26-37 |

So these are the true answers:

| | Your Guess | Correct Answer |
|---|---|---|
|   | 60-80% | 60-80% |
|   | 80-100% | 60-80% |
|   | 20-40% | 0-20% |
|   | 40-60% | 40-60% |
|   | 0-20% | 0-20% |

Now answer the following questions. You can use the information above to decide on the right answer.

These questions test your basic understanding of this task, so **if you answer incorrectly, you may not be eligible for payment**.

Another Applicant F has a score of 21. Would you expect this score to fall within:

- ◯ 20-40%
- ◯ 60-80%

Another Applicant G has a score of 9. Would you expect this score to fall within:

- ◯ 20-40%
- ◯ 0-20%

Applicant H has a score of 35 and Applicant J has a score of 25. Would you expect these scores to fall within:

- ◯ Applicant H to be 20-40% and Applicant J to be 80-100%
- ◯ Applicant H to be 80-100% and Applicant J to be 60-80%

You answered all questions correctly!

In the task we want you to perform now, applicants took a **different** test and **got a score from 0-300**. This score is **unrelated** to the scores in the example (so forget about those!). You will be paid a bonus for being accurate (awarded within 3 days of completing the task). Here are some tips for better accuracy:

1. As you see more scores and get a better understanding of what is a good score or bad score, be flexible and update your previous answers to make them more accurate.
2. Do not rely on intuition or a previous sense of which numbers are “high” or “low.” Scores may not span the entire range from 0 to 300. As in the demo, if you see only small numbers, it is likely that most of the scores in the pool are small.

Keep in mind that you get rewards for correct answers! Good luck!

**Below are scores from 0-300:**

**Number: 262**

Reminder: answering this question accurately will give you a bonus.

| | 0-20% (worst) | 20-40% | 40-60% | 60-80% | 80-100% (best) |
|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ |

**Number: 256**

Reminder: answering this question accurately will give you a bonus.

| | 0-20% (worst) | 20-40% | 40-60% | 60-80% | 80-100% (best) |
|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ |

**Number: 203**

Reminder: answering this question accurately will give you a bonus.

| | 0-20% (worst) | 20-40% | 40-60% | 60-80% | 80-100% (best) |
|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ |

**Number: 237**

Reminder: answering this question accurately will give you a bonus.

| | 0-20% (worst) | 20-40% | 40-60% | 60-80% | 80-100% (best) |
|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ |

**Number: 240**

Reminder: answering this question accurately will give you a bonus.

| | 0-20% (worst) | 20-40% | 40-60% | 60-80% | 80-100% (best) |
|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ |

Feel free to modify your answers if you wish :)

**Number: 275**

Reminder: answering this question accurately will give you a bonus.

| | 0-20% (worst) | 20-40% | 40-60% | 60-80% | 80-100% (best) |
|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ |

**Number: 202**

Reminder: answering this question accurately will give you a bonus.

| | 0-20% (worst) | 20-40% | 40-60% | 60-80% | 80-100% (best) |
|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ |

**Number: 235**

Reminder: answering this question accurately will give you a bonus.

| | 0-20% (worst) | 20-40% | 40-60% | 60-80% | 80-100% (best) |
|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ |

**Number: 216**

Reminder: answering this question accurately will give you a bonus.

| | 0-20% (worst) | 20-40% | 40-60% | 60-80% | 80-100% (best) |
|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ |

**Number: 205**

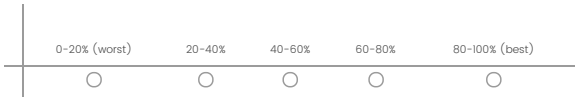
Reminder: answering this question accurately will give you a bonus.

| | 0-20% (worst) | 20-40% | 40-60% | 60-80% | 80-100% (best) |
|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ |

Feel free to modify your answers, including ones from previous pages :)

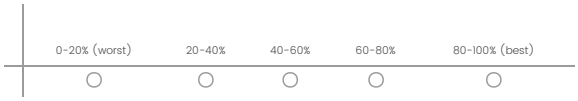
**Number: 238**

Reminder: answering this question accurately will give you a bonus.



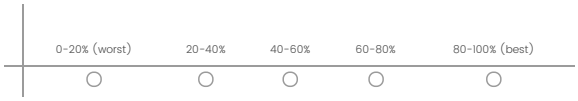

**Number: 251**

Reminder: answering this question accurately will give you a bonus.



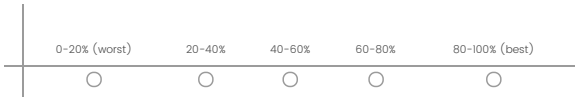

**Number: 251**

Reminder: answering this question accurately will give you a bonus.



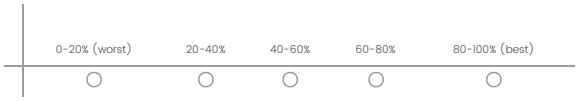

**Number: 203**

Reminder: answering this question accurately will give you a bonus.




**Number: 228**

Reminder: answering this question accurately will give you a bonus.




Feel free to modify your answers, including ones from previous pages :)

**Number: 204**

Reminder: answering this question accurately will give you a bonus.

| | 0-20% (worst) | 20-40% | 40-60% | 60-80% | 80-100% (best) |
|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ |

**Number: 219**

Reminder: answering this question accurately will give you a bonus.

| | 0-20% (worst) | 20-40% | 40-60% | 60-80% | 80-100% (best) |
|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ |

**Number: 216**

Reminder: answering this question accurately will give you a bonus.

| | 0-20% (worst) | 20-40% | 40-60% | 60-80% | 80-100% (best) |
|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ |

**Number: 217**

Reminder: answering this question accurately will give you a bonus.

| | 0-20% (worst) | 20-40% | 40-60% | 60-80% | 80-100% (best) |
|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ |

**Number: 224**

Reminder: answering this question accurately will give you a bonus.

| | 0-20% (worst) | 20-40% | 40-60% | 60-80% | 80-100% (best) |
|---|---|---|---|---|---|
| | ○ | ○ | ○ | ○ | ○ |

Feel free to modify your answers, including ones from previous pages :)

## Debriefing

Thank you for your participation!

We would like to let you know that in this study we are collecting data about errors individuals make when reviewing job applicants. A common problem in human evaluations is being too tough or too lenient, so applicants are often hired or rejected depending mainly on the strictness of the evaluator. This reviewer bias can result in the hiring of the wrong candidate or the exclusion of good candidates.

With this particular survey, we're investigating how evaluating more applications can improve a reviewer's judgment. Essentially, does giving reviewers more applications to look at give them a better sense of how good each application is compared to all the rest?

Using the data collected here, we will come up with strategies to reduce the impact of human bias and error in evaluating applications.

If you have any questions about this study, please contact Jingyan Wang at jingyanw@cs.cmu.edu.

If you have any comments, or suggestions on how to improve this task, please leave them here:

Here is your survey completion code: 651F86E5

Copy this value to paste into Prolific. Then click next and proceed to the next page to finish the survey.

We thank you for your time spent taking this survey.
Your response has been recorded.