Your job is simple: try to reproduce the results of the student you’ve been assigned to (“the author”). Do your best to reproduce the results on your own, but you may contact the author if absolutely necessary. If they make any changes in order to enable you to reproduce their results, they must check these changes into their git repository.

1. Briefly describe the steps you had to follow in order to reproduce the author’s results.
   
   (a) How long did it take you?
   
   (b) Did you have to do anything not documented in the README / report?

2. Did the results you got from reproducing the author’s results match what they put into their report?
   
   (a) Did you have to try more than once, and/or contact the author in order to get the results to match?
   
   (b) If they didn’t match in the end, why do you think that might be?

3. Did the author’s results and conclusion match your own?
   
   (a) If so, does this increase your confidence in your own assessment?
   
   (b) If not, which result are you more confident in?

4. How well did the author use the evaluation techniques covered in this class?
   
   (a) Did they use an appropriate method to select the number of runs per experiment?

   (b) When reporting results, did they indicate what type of index of central tendency (eg. mean, median, mode) they used, and if appropriate, did they justify why?

   (c) Did they present some measure of variability? (eg. error bars in plots, confidence intervals, reporting variance or stddev, etc.) If presented, did they justify the one they used?

   (d) If mean or median results were similar between the alternatives, did the author use appropriate techniques (eg. examining confidence intervals for the means) to show that the results were or were not statistically significant?

   (e) Did you notice any of the common mistakes we’ve discussed (eg. the ones in Chapter 2 from the book)?