

Review export of

FIN: First Report

Reviews written by

Anonymous user

Anonymous user

Anonymous user

Anonymous user

Anonymous user



Downloaded at 29 Jan 2025,
17:38



Review for O.J.B. Timmers (Ole)

Accessibility consideration

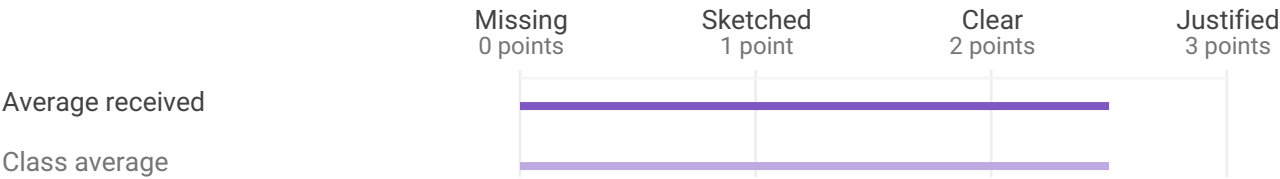
This file does not fully conform with all applicable guidelines for accessible digital documents

[OPEN ALTERNATIVE](#)

Rubric section

First Metric Intuition

Rubric criterion



Received ratings

- Justified - The intuitive explanation present a causal explanantion on why the intervention should work and produce a desired effect on the proposed metrics to be measured.
Anonymous user, Anonymous user

Clear - The explanation is consistent and do not present aspects of reverse causation. However the explanation does not really explain why the intervention would work.
Anonymous user, Anonymous user

4 comments



Anonymous user
24-01-2025 23:20:57

The intuition behind the SumCorrectness metric is straightforward, aiming to measure how accurately participants identify vulnerabilities

1 reply



O.J.B. Timmers (Ole)
29-01-2025 14:48:12

Thank you!



Anonymous user
24-01-2025 05:09:27

The intuition behind the first metric is fairly justified



Anonymous user
24-01-2025 18:54:27

[@page 2](#)

Good explanation

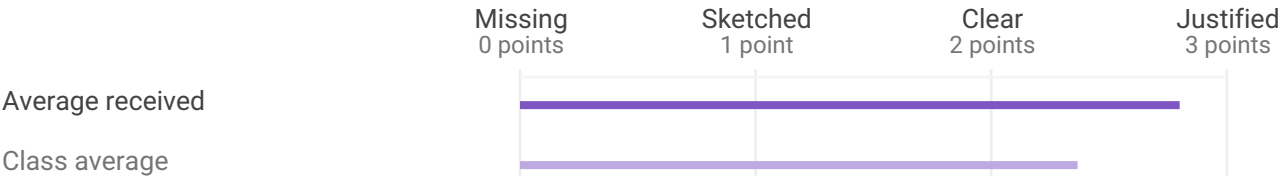


Anonymous user
24-01-2025 01:05:54

Makes sense for the intervention

Second Metric Intuition

Rubric criterion



Received ratings

- Justified - The intuitive explanation present a causal explanantion on why the intervention should work and produce a desired effect on the proposed metrics to be measured.
Anonymous user, Anonymous user, Anonymous user

Clear - The explanation is consistent and do not present aspects of reverse causation. However the explanation does not really explain why the intervention would work.
Anonymous user

4 comments



Anonymous user
24-01-2025 23:21:20

: The intuition for SumDifficulty is clear but not fully justified. While it evaluates the perceived difficulty of finding vulnerabilities, the report could better explain why this is an important success indicator beyond participant ease of use.

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:08:42

Thanks, I'll explain better why the metric is an important success indicator.



Anonymous user
24-01-2025 05:14:12

The intuition seems to be consistent with the intended intervention goals, no issues here



Anonymous user
24-01-2025 18:55:54

[@page 2](#)

This is also a justified selection but it is not really derived from the intuition described here. Maybe the phrasing could be improved to focus on the actual improvement that would be derived for the second metric (i.e. ease of use)

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:11:00

Thanks, I'll try to focus it more on the actual improvement at the second metric.

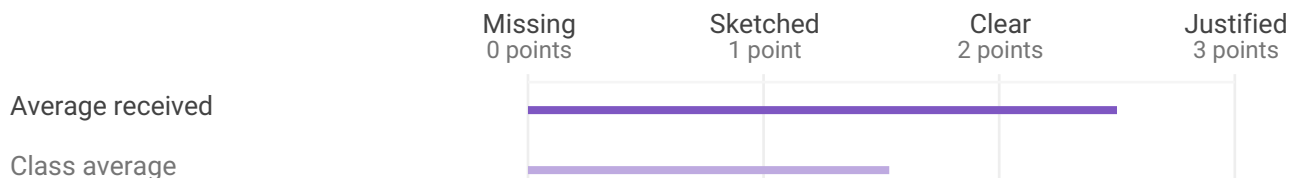


Anonymous user
24-01-2025 01:05:59

Makes sense for the intervention

Figure with Causal Links

Rubric criterion



Received ratings

Justified - The figure presents causal relations among the main aspects of the problem. Every relation seems justified.

Anonymous user, Anonymous user

Clear - Most relations seems to be causal, albeit some of the relations seems a bit arbitrary and not justified.

Anonymous user, Anonymous user

4 comments



Anonymous user

24-01-2025 23:22:16

The relationships between metrics, demographics, and intervention impact are outlined and well justified



Anonymous user

24-01-2025 05:15:19

Very comprehensive diagram, but can be simplified a bit

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:12:46

Thanks, but I think I'll let it this way because it states everything needed to know.



Anonymous user
24-01-2025 18:57:14

Given that this is a causal figure, you should point the participants data to the cleaned up data.

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:11:54

Sharp, will edit this.



Anonymous user
24-01-2025 01:06:39

Extremely detailed, text could be a little larger

1 reply

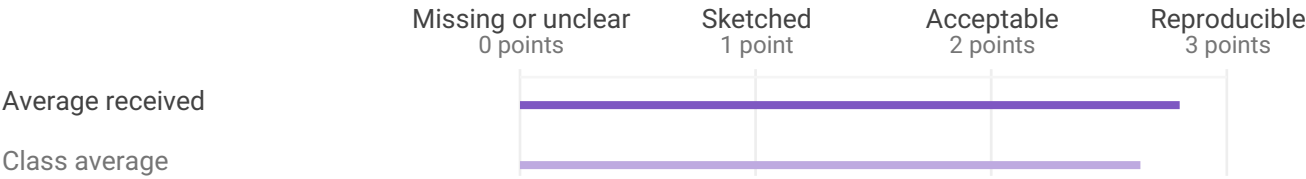


O.J.B. Timmers (Ole)
29-01-2025 15:13:36

Thanks, I'll see if I can get it bigger.

Variable definition 1st metric

Rubric criterion



Received ratings

- Reproducible - The high level idea is clear and enough details are given. The variable present in the formula are essentially clear. Given the raw data you would be able able to reconstruct them.
Anonymous user, Anonymous user, Anonymous user
- Acceptable - The high level idea is clear and some descriptions on how this can be calculated is present. Not all variables presented in the formula are clear. You would not be able to reconstruct them without talking to the author.
Anonymous user

4 comments



Anonymous user
24-01-2025 23:22:46

Variables for SumCorrectness, including the definitions of participant-identified and actual vulnerable lines, are well-defined. The formula is straightforward and easy to interpret.

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:13:49

Thanks!



Anonymous user
24-01-2025 05:16:08

Fairly reproducible, all variables required are defined

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:14:17

Thanks!



Anonymous user
24-01-2025 18:57:53

I like the mathematical presentation if the variables and the explanation for the division.

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:14:08

Thanks!



Anonymous user
24-01-2025 01:08:10

Well defined

1 reply

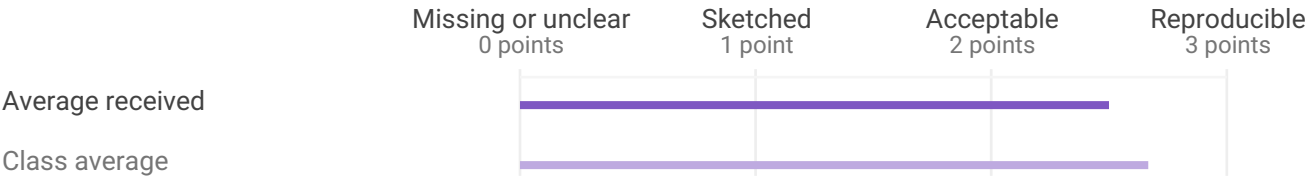


O.J.B. Timmers (Ole)
29-01-2025 15:14:30

Thanks!

Formula 1st Metric

Rubric criterion



Received ratings

Reproducible - The formula is well described, you would be able to implement it in Excel and it is clear that the metrics corresponds to the text.
Anonymous user, Anonymous user

Acceptable - The formula is well described and you would be able to implement it in Excel with some explanantions from the authors. However, it is unclear whether the formual actually correspond to the metric of success.
Anonymous user, Anonymous user

4 comments



Anonymous user
24-01-2025 23:23:03

The formula for calculating SumCorrectness is correctly specified. Including examples of scenarios where participants perform well versus poorly strengthens the clarity of this section.



Anonymous user
24-01-2025 05:17:44

Formula seems to be a little incomplete, for example, what was the process to calculate Ci?

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:17:02

The variable explanation of C_i states this, but I'll look if can make this more clear. i.e.: I'll change the explanation to: The amount of lines from the participant *that are* equal with the actual lines of the vulnerability.



Anonymous user
24-01-2025 18:58:37

The formula is straight forward and pleasantly presented

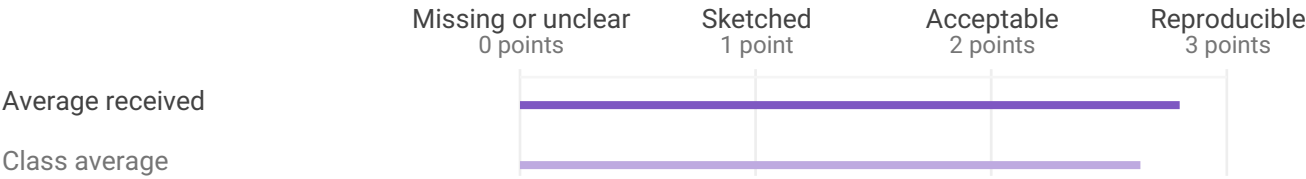


Anonymous user
24-01-2025 01:08:33

Formula is clear and usable

Direction 1st Metric

Rubric criterion



Received ratings

Reproducible - The direction of success is explained for all variable and it is clear in which direction we have an increase and why such a change would derive from the changes in the input variable.
Anonymous user, Anonymous user, Anonymous user

Acceptable - The direction of success is explained for all variable and it is clear in which direction we have a success. It is still unclear why such a change would derive from the changes in the input variable.
Anonymous user

4 comments



Anonymous user
24-01-2025 23:23:30

The desired direction (higher SumCorrectness is better) is mentioned but lacks thresholds for what constitutes "good" or "poor" performance. Providing benchmarks would enhance interpretability.

2 replies



O.J.B. Timmers (Ole)
29-01-2025 15:20:27

Thanks, and I agree somewhat on the insight you give now, however, I'm going to leave it like this:
Because I state that one category to the other should be higher by average, where as of statistics by common, it should be **significantly higher** - 5%.



O.J.B. Timmers (Ole)
29-01-2025 15:21:25

Making that, there is no need here to know what good or poor performances are; one should just be higher than the other.



Anonymous user
24-01-2025 05:18:35

Direction makes sense and is explained well



Anonymous user
24-01-2025 18:59:30

This is a good explanation of the metric as well.

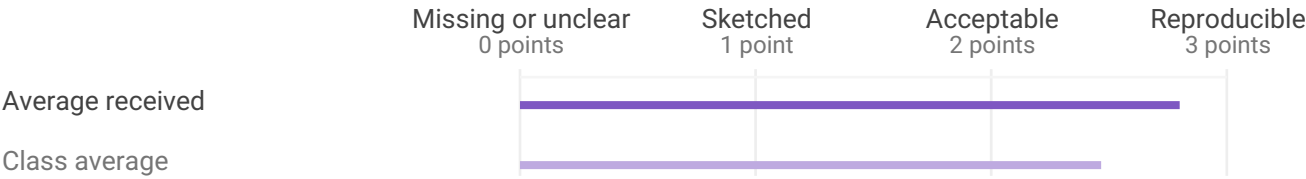


Anonymous user
24-01-2025 01:08:44

Direction is clear for the intervention

Justification 1st Metric

Rubric criterion



Received ratings

Reproducible - It is explained why the intervention would impact some of the input variables, and it is clear and consistent that such a changewoud generate a better success (i.e. lower or higher final metric, see previous point).
Anonymous user, Anonymous user, Anonymous user

Acceptable - It is explained why the intervention would impact some of the input variables, but it is not yet explained why this change in the variable woud generate a better success (i.e. lower or higher final metric, see above point).
Anonymous user

4 comments



Anonymous user
24-01-2025 23:23:52

The report states that higher SumCorrectness indicates successful intervention but does not discuss potential limitations, such as how scenario complexity might influence scores independently of the intervention.

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:25:22

Good one, will look into that.



Anonymous user
24-01-2025 05:19:18

Justification is clear and concise



Anonymous user
24-01-2025 19:00:36

@page 3

The explanation is straight forward generally but I dont really get what this means. What is the grade? Do you mean the rate?

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:26:40

Yeah you're right lol, don't why I wrote *grade* here - inconsistent. Thanks

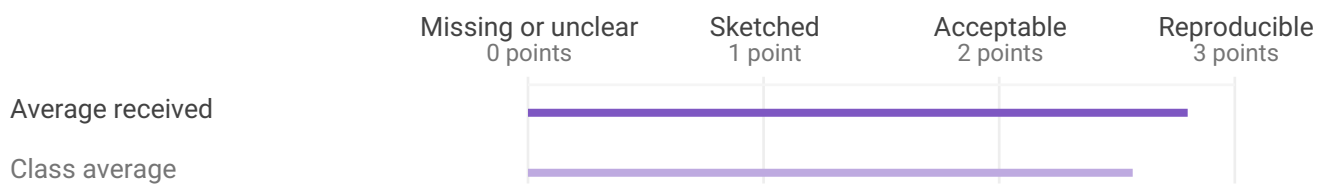


Anonymous user
24-01-2025 01:08:50

Justified

Variable definition 2nd Metric

Rubric criterion



Received ratings

Reproducible - The high level idea is clear and enough details are given. The variable present in the formula are essentially clear. Given the raw data you would be able able to reconstruct them.
Anonymous user, Anonymous user, Anonymous user

Acceptable - The high level idea is clear and some descriptions on how this can be calculated is present. Not all variables presented in the formula are clear. You would not be able to reconstruct them without talking to the author.
Anonymous user

4 comments



Anonymous user
24-01-2025 23:24:13

The variables for SumDifficulty, including the transformation of difficulty scores into a consistent scale, are well-defined and easy to understand.



Anonymous user
24-01-2025 05:20:13

Variables are well defined, easy to understand



Anonymous user
24-01-2025 19:00:53

Same as the first metric, well presented

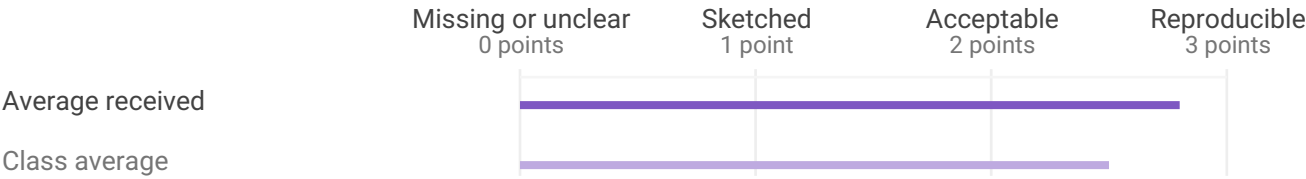


Anonymous user
24-01-2025 01:11:37

Clear

Formula 2nd Metric

Rubric criterion



Received ratings

- Reproducible - The formula is well described, you would be able to implement it in Excel and it is clear that the metrics corresponds to the text.
Anonymous user, Anonymous user, Anonymous user

Acceptable - The formula is well described and you would be able to implement it in Excel with some explanantions from the authors. However, it is unclear whether the formual actually correspond to the metric of success.
Anonymous user

4 comments

- Anonymous user
24-01-2025 23:24:34

The formula for SumDifficulty is precise and includes all relevant components. The summation approach aligns well with the study's objectives.
- Anonymous user
24-01-2025 05:20:57

Formula corresponds to the metric well, easily reproducible
- Anonymous user
24-01-2025 19:03:09

Same as the first metric, well presented.

Now that I think about it, maybe it would be better to show the second metric as an average instead of the sum. It would keep it balanced.

However, given that the experiment has an equal distribution of participant cases with slice/full, this wont make a difference other than aesthetic

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:30:28

Yeah you point out a good point, I had averages at first, but then Aurora said the amount of sliced/full types of code are always the same distributed by definition. That way it is more logically to do a summation instead of doing a average. - i.e. I'll leave it like that but good point you point out.



Anonymous user
24-01-2025 01:11:41

I like how you separated the "slice" and "full" types for the metric (this goes for M1 as well)

1 reply

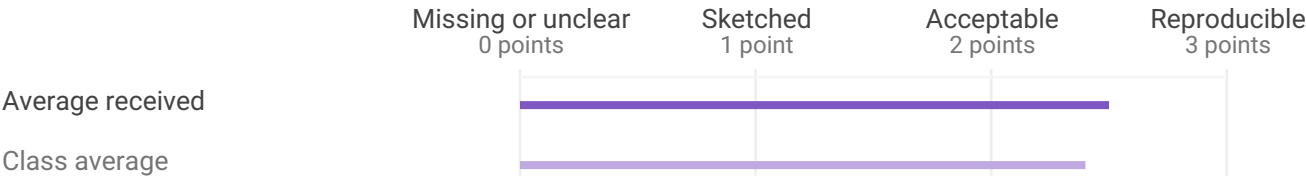


O.J.B. Timmers (Ole)
29-01-2025 15:31:20

Thanks!

Direction 2nd Metric

Rubric criterion



Received ratings

Reproducible - The direction of success is explained for all variable and it is clear in which direction we have an increase and why such a change would derive from the changes in the input variable.
Anonymous user, Anonymous user

Acceptable - The direction of success is explained for all variable and it is clear in which direction we have a success. It is still unclear why such a change would derive from the changes in the input variable.
Anonymous user, Anonymous user

4 comments



Anonymous user
24-01-2025 23:24:56

The direction (lower SumDifficulty is better) is stated but lacks context, such as how much lower difficulty is considered meaningful or impactful. Adding this detail would strengthen the metric's utility.

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:31:46

I'll add some more context here, thanks.



Anonymous user
24-01-2025 05:22:24

Since the direction seems to be a comparison between 2 end values, maybe it can be converted into a ratio between the 2 and then defined to be higher or lower? But still works as it is

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:39:21

I'm not sure what adaptation is being meant here.



Anonymous user
24-01-2025 19:03:57

Straight forward, nothing to add

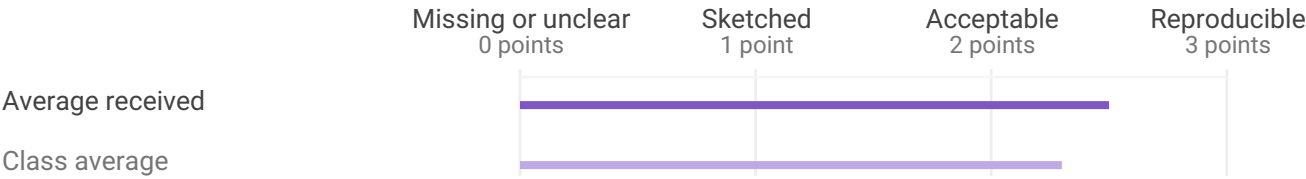


Anonymous user
24-01-2025 01:12:01

Direction makes sense

Justification 2nd Metric

Rubric criterion



Received ratings

Reproducible - It is explained why the intervention would impact some of the input variables, and it is clear and consistent that such a changewoud generate a better success (i.e. lower or higher final metric, see previous point).
Anonymous user, Anonymous user

Acceptable - It is explained why the intervention would impact some of the input variables, but it is not yet explained why this change in the variable woud generate a better success (i.e. lower or higher final metric, see above point).
Anonymous user, Anonymous user

4 comments



Anonymous user
24-01-2025 23:25:22

The justification for SumDifficulty is logical but limited. While reducing difficulty is desirable, the report does not explore how this reduction might translate into improved participant performance or outcomes.

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:40:01

I'll add more explanation here, thanks!



Anonymous user
24-01-2025 05:23:29

Well explained, but maybe a more elaborate explanation can help as difficulty is more of a perception than a calculated metric

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:40:36

I'll add some more explanation, thanks!



Anonymous user
24-01-2025 19:04:13

also straight forward.

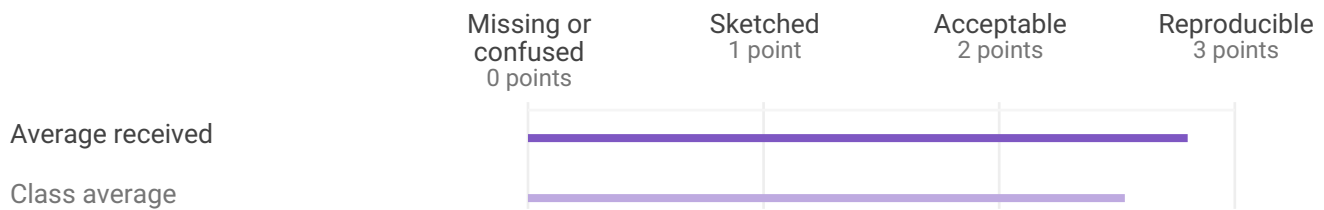


Anonymous user
24-01-2025 01:12:06

Justified

Algorithm 1st Metric

Rubric criterion



Received ratings

Reproducible - You will be essentially be able to reproduce the steps yourself given the raw dat csv. You will not have difficulties in understanding which variables of the raw data are used to perform each step..
Anonymous user, Anonymous user, Anonymous user

Acceptable - There is a list of the steps that makes a clear reference to the raw data. Albeit the data is not excatly clear you would be able to find out with some trial and error. some transformation of the data would be a bit unclear.
Anonymous user

4 comments



Anonymous user
24-01-2025 23:25:47

The steps for calculating SumCorrectness are logically structured, from normalizing participant responses to aggregating scores. The detailed explanation ensures reproducibility.



Anonymous user
24-01-2025 05:24:07

Algorithm is well written, easy to follow



Anonymous user
24-01-2025 19:05:05

@page 4

I appreciate the addition of data cleaning and parsing from the actual raw data instead of assuming that data was cleaned before this algorithm

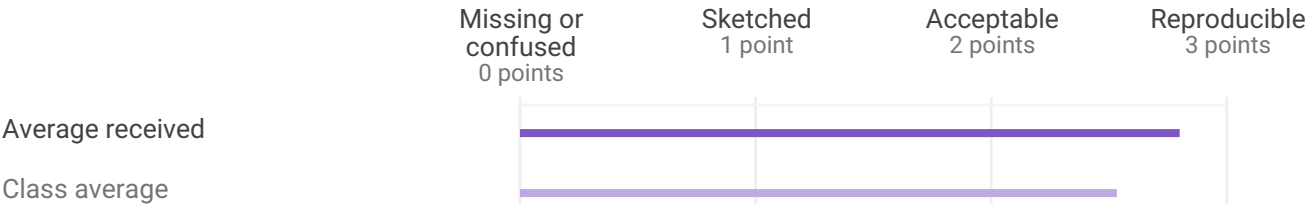


Anonymous user
24-01-2025 01:12:27

Detailed and followable

First Example 1st Metric

Rubric criterion



Received ratings

Reproducible - The first example explains well what the participant has and why this can be considered a bad or a good result,, i.e. from this participant it seems that s/he managed to perform the tasks. (as if this was failure, or as if this was a success).
Anonymous user, Anonymous user, Anonymous user

Acceptable - The first example explains the key ideas of what the participant has done. It is a still a bit unclear whether this is a bad or a good result from the perspective of the intervention.
Anonymous user

4 comments



Anonymous user
24-01-2025 23:26:05

the example illustrates a positive outcome where participants perform better with sliced code. This effectively demonstrates how the metric captures improvements due to the intervention.



Anonymous user
24-01-2025 05:25:08

It is a good example, but needs to be more specific with the calculations (can be shown by substituting values into the formula and deriving end results)

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:43:10

Good feedback, but I'll think about this, I wanted to let the examples be compact here.



Anonymous user
24-01-2025 19:05:57

Maybe it would be nice to show what C and L actually are for the example in order to justify the computation

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:42:10

I'm not sure what is being meant with C and L.

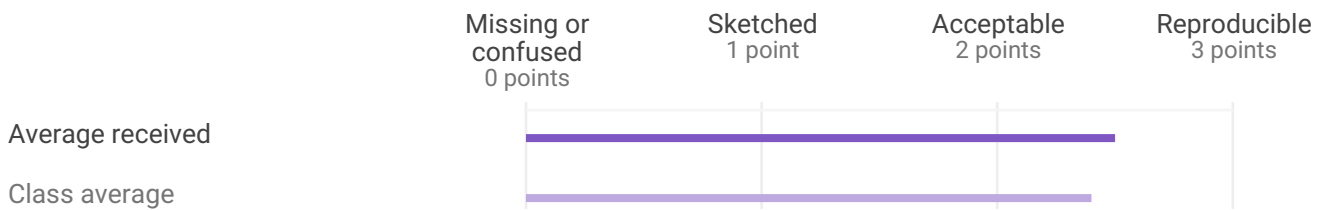


Anonymous user
24-01-2025 01:13:02

Great example

Second Example 1st Metric

Rubric criterion



Received ratings

Reproducible - The first example explains well what the participant has and why this can be considered a bad or a good result, i.e. from this participant it seems that s/he managed to perform the tasks (as if this was failure, or as if this was a success)

Anonymous user, Anonymous user

Acceptable - The second example explains the key ideas of what the participant has done. It is still a bit unclear whether this is a bad or a good result from the perspective of the intervention.

Anonymous user, Anonymous user

4 comments



Anonymous user

24-01-2025 23:26:26

A contrasting example highlights a scenario where the intervention does not improve performance. This balance in examples strengthens the section.



Anonymous user

24-01-2025 05:25:28

Similar to the previous example



Anonymous user

24-01-2025 19:06:03

Maybe it would be nice to show what C and L actually are for the example in order to justify the computation

1 reply



O.J.B. Timmers (Ole)
29-01-2025 15:43:37

Same as previous, not sure what is being meant here.

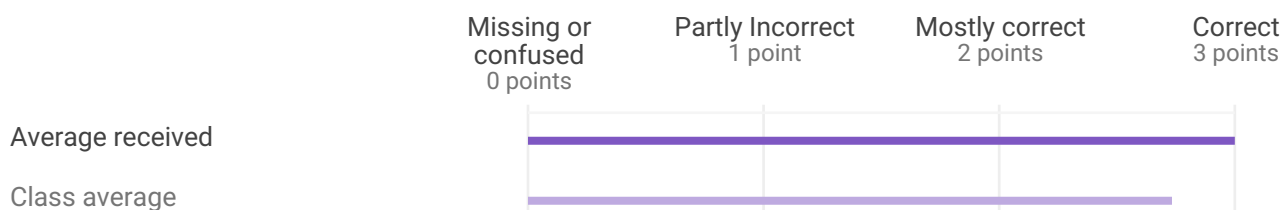


Anonymous user
24-01-2025 01:13:08

Another one

Correctness 1st Metric

Rubric criterion



Received ratings

Correct - All steps seem correct assuming that they would be correctly implemented in the code. They seem to fully correspond at the intended calculation proposed metrics of success
Anonymous user, Anonymous user, Anonymous user, Anonymous user

4 comments



Anonymous user
24-01-2025 23:26:40

The metric is correctly implemented and validated with examples. The logic aligns with the goals of the experiment.



Anonymous user
24-01-2025 05:26:25

Overall the proposed algorithm seems to work very well, and fits well with the goal of the intervention



Anonymous user
24-01-2025 19:06:16

The metric seems to be computed as expected.

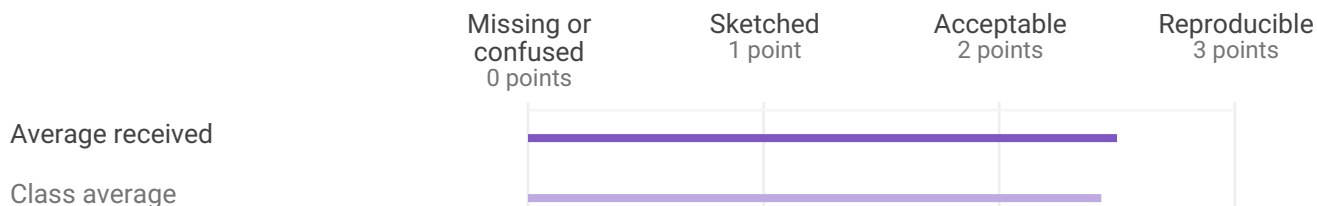


Anonymous user
24-01-2025 01:13:15

Seems correct

Algorithm 2nd Metric

Rubric criterion



Received ratings

Reproducible - You will be essentially be able to reproduce the steps yourself given the raw dat csv. You will not have difficulties in understanding which variables of the raw data are used to perform each step..
Anonymous user, Anonymous user

Acceptable - There is a list of the steps that makes a clear reference to the raw data. Albeit the data is not excatly clear you would be able to find out with some trial and error. some transformation of the data would be a bit unclear.
Anonymous user, Anonymous user

4 comments



Anonymous user
24-01-2025 23:26:54

The steps for calculating SumDifficulty are clear, with a consistent approach for scaling and aggregation. The explanation is comprehensive and ensures transparency.



Anonymous user
24-01-2025 05:27:45

well defined algorithm, could be a bit more specific regarding the the ratings of the difficulty scores

1 reply



O.J.B. Timmers (Ole)
29-01-2025 16:24:21

Will look into that!



Anonymous user
24-01-2025 19:07:04

Good explanation of the process.

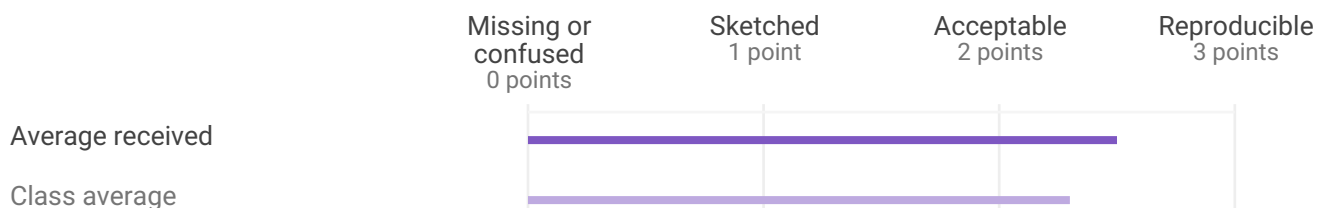


Anonymous user
24-01-2025 01:12:40

Detailed and followable

First Example 2n Metric

Rubric criterion



Received ratings

Reproducible - The first example explains well what the participant has and why this can be considered a bad or a good result,, i.e. from this participant it seems that s/he managed to perform the tasks. (as if this was failure, or as if this was a success).

Anonymous user, Anonymous user

Acceptable - The first example explains the key ideas of what the participant has done. It is a still a bit unclear whether this is a bad or a good result from the perspective of the intervention.

Anonymous user, Anonymous user

4 comments



Anonymous user

24-01-2025 23:27:13

The example demonstrates how sliced code leads to lower difficulty scores, illustrating the intervention's intended effect.



Anonymous user

24-01-2025 05:28:27

Again, good example, being more specific and implementing the example using the formula can be immensely helpful



Anonymous user

24-01-2025 19:08:17

Good example and good explanation.



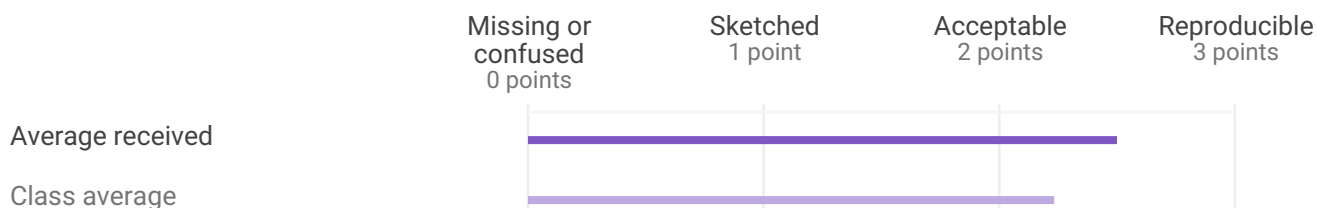
Anonymous user

24-01-2025 01:13:29

Wonderful example

Second Example 2nd Metric

Rubric criterion



Received ratings

Reproducible - The first example explains well what the participant has and why this can be considered a bad or a good result, i.e. from this participant it seems that s/he managed to perform the tasks (as if this was failure, or as if this was a success)

Anonymous user, Anonymous user

Acceptable - The second example explains the key ideas of what the participant has done. It is still a bit unclear whether this is a bad or a good result from the perspective of the intervention.

Anonymous user, Anonymous user

4 comments



Anonymous user

24-01-2025 23:27:30

The example of higher difficulty scores with full code provides a useful contrast, reinforcing the metric's purpose and sensitivity.



Anonymous user

24-01-2025 05:28:43

Similar to the previous example



Anonymous user

24-01-2025 19:08:20

Good example and good explanation.



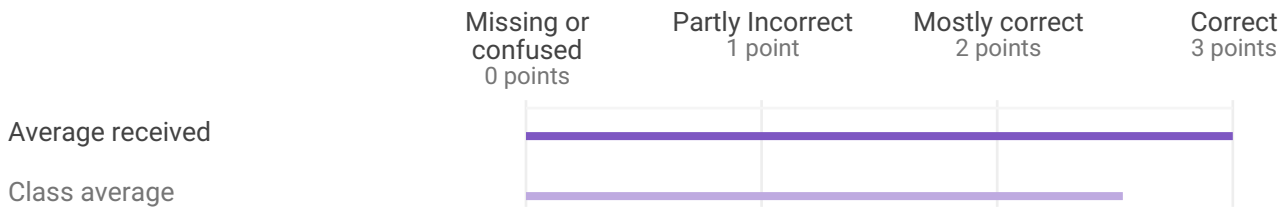
Anonymous user

24-01-2025 01:13:41

Once again, spectacular

Correctness 2nd Metric

Rubric criterion



Received ratings

Correct - All steps seem correct assuming that they would be correctly implemented in the code. They seem to fully correspond at the intended calculation proposed metrics of success
Anonymous user, Anonymous user, Anonymous user, Anonymous user

4 comments



Anonymous user
24-01-2025 23:27:46

The computation of SumDifficulty is correct, with examples validating its logical design and alignment with the intervention's goal



Anonymous user
24-01-2025 05:30:03

Overall, this algorithm aligns well with the goals of the intervention and corresponds at the intended calculation proposed metrics of success



Anonymous user
24-01-2025 19:10:19

Maybe it would be nice to transform the second metric and the first metric into a value of 1 and -1 where:

If Metric Slice > Metric Full => M1 = 1, M2 = -1.

And the opposite order for the other case.

This would be just so that the data expected for the coolab notebook to be in the right format

1 reply



O.J.B. Timmers (Ole)
29-01-2025 16:26:34

I understand you, but I changed the colab to let it work with my metrics and that worked out good.

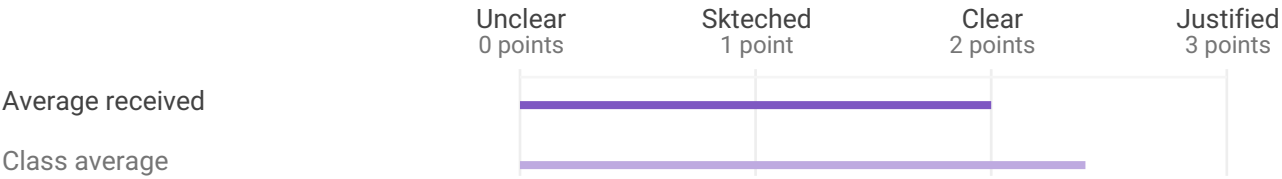


Anonymous user
24-01-2025 01:13:58

Seems correct and viable

Demographics Variables

Rubric criterion



Received ratings

Justified - Both variables are described and you would be able to find them in the raw data. It is also clear why these variables have been chosen and how they will be finally used.
Anonymous user

Clear - Both variables are present and you would be able to find them in the raw data. It is unclear why these variables have been chosen from the perspective of the intervention.
Anonymous user, Anonymous user

Sketched - The demographics variables are partly described (e.g. there is only one variable). You will have difficulty in finding them in the raw data without a bit of trial and error.
Anonymous user

4 comments



Anonymous user
24-01-2025 23:28:08

The demographic variables (coding and vulnerability detection experience) are clearly defined and appropriately scaled. The analysis identifies trends and their potential impact on outcomes.



Anonymous user
24-01-2025 05:30:42

could be explained in more detail

1 reply



O.J.B. Timmers (Ole)
29-01-2025 16:30:22

Will do



Anonymous user
24-01-2025 19:11:29

@page 4

But in what way will it give a better understanding?

1 reply



O.J.B. Timmers (Ole)
29-01-2025 16:30:14

Sharp, will elaborate more.



Anonymous user
24-01-2025 01:15:05

Description of variables is very vague, could've included the source of data and method of calculation

1 reply

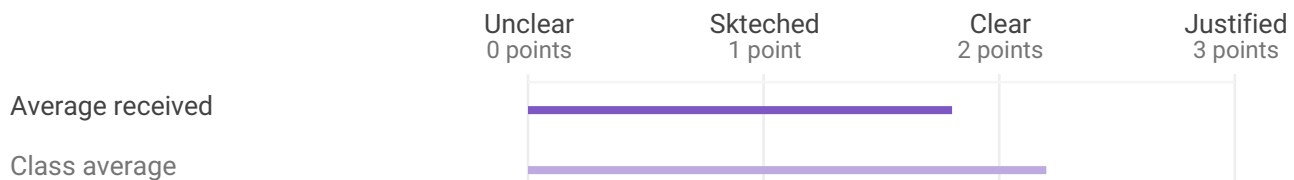


O.J.B. Timmers (Ole)
29-01-2025 16:30:59

I'll elaborate more where needed.

Perception Variable

Rubric criterion



Received ratings

Clear - The variable is present and you would be able to find it in the raw data. It is bit unclear why this variables has been chosen from the perspective of the intervention.
Anonymous user, Anonymous user, Anonymous user

Sketched - The perception variable is partly described. You will have difficulty in finding it in the raw data without a bit of trial and error.
Anonymous user

4 comments



Anonymous user
24-01-2025 23:29:11

The confidence variable is defined and deeply analyzed.



Anonymous user
24-01-2025 05:31:39

it is well explained, easy to understand



Anonymous user
24-01-2025 19:11:53

You should explain why this variable was chosen

1 reply



O.J.B. Timmers (Ole)
29-01-2025 16:31:52

Will do.



Anonymous user
24-01-2025 01:16:10

A reference to the raw data would be nice

1 reply

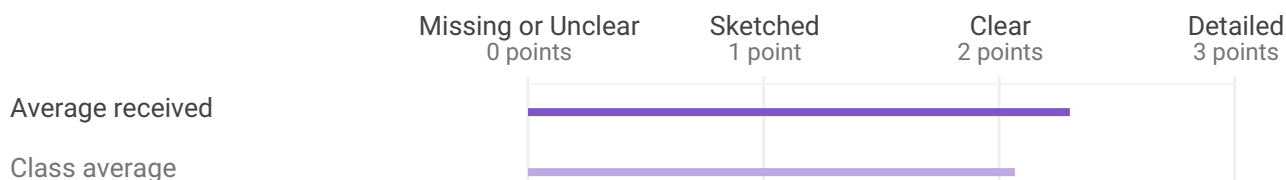


O.J.B. Timmers (Ole)
29-01-2025 16:31:45

I'll refer to it.

Distribution of Demographics variables

Rubric criterion



Received ratings

Detailed - Text and figure are coherent and the text explains whether this information is important for the intervention.

Anonymous user, Anonymous user

Clear - The text and the figure are coherent with one another but it is not explained whether this phenomenon is important or not.

Anonymous user

Sketched - The text and the figure are incoherent with one another (e.g there is a peak at point and the text says that all values are equally present or viceversa)

Anonymous user

4 comments



Anonymous user

24-01-2025 23:30:04

The report provides clear histograms and scatter plots illustrating participant distributions. The analysis identifies key trends, such as the dominance of novice participants.



Anonymous user

24-01-2025 05:32:27

Diagrams and text are consistent, no issues regarding this part



Anonymous user

24-01-2025 19:13:33

@page 6

The diagram caption says something different than the text + what does it mean for it to be inappropriate or appropriate for the intervention in this case?

1 reply



O.J.B. Timmers (Ole)
29-01-2025 16:37:13

Good feedback, with the text was meant that the group of experienced **are** appropriate for the intervention, I should make that more clear. For the rest it should be inline with the descriptions. I'll look into it.

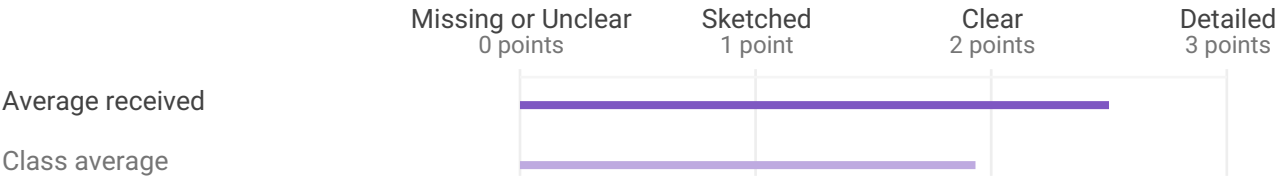


Anonymous user
24-01-2025 01:03:30

Explains relation to the intervention well, great stuff

Scatter Plot of Demographics Variables

Rubric criterion



Received ratings

- Detailed - Text and figure are coherent and the test explains whether this information is important for the intervention.
Anonymous user, Anonymous user

Clear - The test and the figure are coherent with one another but it is not explained whether this phenomenon is important or not.
Anonymous user, Anonymous user

4 comments



Anonymous user
24-01-2025 23:30:22

The scatter plot effectively visualizes the relationship between coding and vulnerability detection experience but lacks discussion of notable patterns or outliers.

1 reply



O.J.B. Timmers (Ole)
29-01-2025 16:37:46

I'll look into it.



Anonymous user
24-01-2025 05:33:20

A more detailed explanation regarding the importance of this would be helpful

1 reply



O.J.B. Timmers (Ole)
29-01-2025 16:38:36

I'm not sure what is being meant with the importance of ...?



Anonymous user
24-01-2025 19:15:11

[@page 6](#)

This type of explanation is good and should be there for all cases where the appropriateness is addressed.

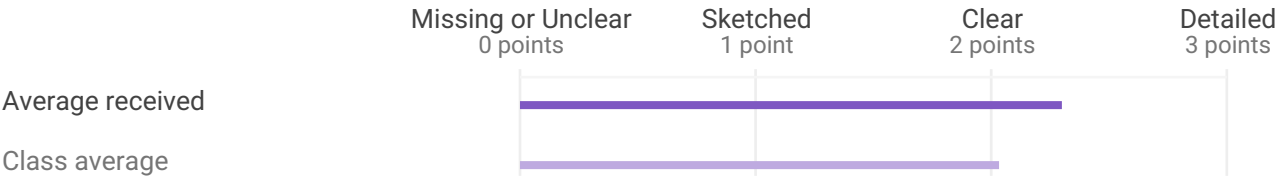


Anonymous user
24-01-2025 01:03:03

Very detailed and informative

Impact on the intervention

Rubric criterion



Received ratings

- Detailed - The final text summarizes the findings and draw conclusion by pointing to some criticalities (o stating that there are no criticalities) from the perspective of the intervention.

Anonymous user
- Clear - The final text summarizes the findings but doesn't draw conclusion or pointing to criticalities (o stating that there are no criticalities).

Anonymous user, Anonymous user, Anonymous user

5 comments



Anonymous user
24-01-2025 23:30:36

The intervention is shown to improve correctness and reduce difficulty, supported by descriptive statistics and box plots. This analysis is well-articulated.



Anonymous user
24-01-2025 05:34:47

It is summarized well, but missing the key criticalities and conclusions



Anonymous user
24-01-2025 19:17:10

@page 6

This is a good conclusion based on the table data



Anonymous user
24-01-2025 19:17:48

@page 6

Maybe rephrase this for better readability?

1 reply



O.J.B. Timmers (Ole)
29-01-2025 17:22:01

Thanks, will check.

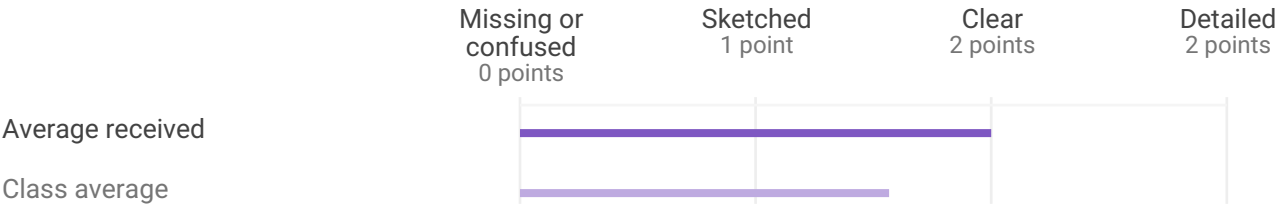


Anonymous user
24-01-2025 01:17:25

Impact on the intervention is explained

Distribution of the intervention

Rubric criterion



Received ratings

Detailed - All tables and figures are present and you would be able to compare the results with your results or those of your colleagues. However it is unclear whether this a positive or successful intervention or whether there where problems with the scenarios
Anonymous user, Anonymous user, Anonymous user, Anonymous user

Clear - All tables, figures, and the corresponding texts are present and you would be able to compare the results with your results or those of your colleagues. However it is unclear whether this a positive or successful intervention or whether there where problems with the scenarios
Anonymous user, Anonymous user, Anonymous user, Anonymous user

4 comments



Anonymous user
24-01-2025 23:30:56

The distribution of metrics across groups is effectively represented. The analysis highlights key differences between sliced and full code.



Anonymous user
24-01-2025 05:36:51

While most diagrams are present, the final conclusion seems to be hazy

1 reply



O.J.B. Timmers (Ole)
29-01-2025 17:28:09

Will check on that, thanks!



Anonymous user
24-01-2025 19:19:34

[@page 6](#)

It is not really evenly distributed. You found something that might be relevant ("DoS-vulnerabilities were almost never found") this is something to mention either way.

1 reply



O.J.B. Timmers (Ole)
29-01-2025 17:29:00

Good point, will mention that point.

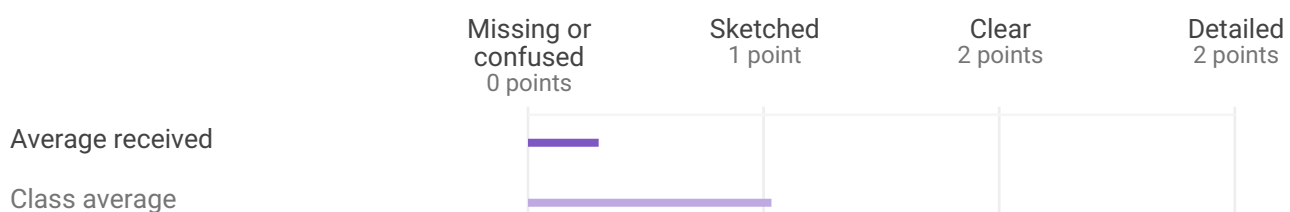


Anonymous user
24-01-2025 01:17:37

Results are clear

Statistics of Intervention

Rubric criterion



Received ratings

Sketched - Some tables and figures are present but you would not be able to compare the results with your results or those of your colleagues. For example you would find the intervention but not the scenarios
Anonymous user

Missing or confused - Most figures/tables are missing or confused. You would not be able to explain even the high level finding of the presented results.
Anonymous user, Anonymous user, Anonymous user

4 comments



Anonymous user
24-01-2025 23:32:30

does not exist



Anonymous user
24-01-2025 05:37:21

missing



Anonymous user
24-01-2025 19:19:53

@page 6

Make sure to add it for the final submission

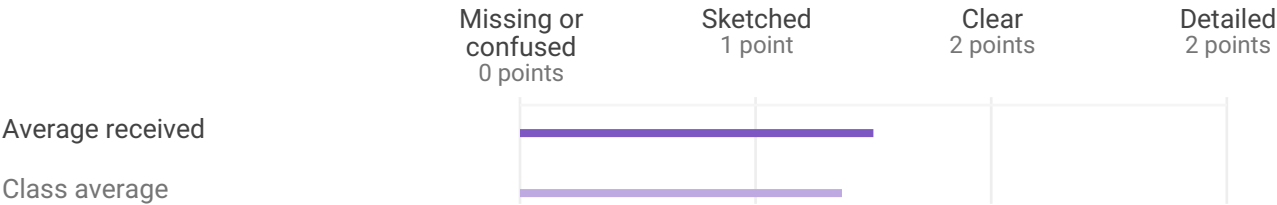


Anonymous user
24-01-2025 01:02:16

Missing

Summary

Rubric criterion



Received ratings

- Detailed - The summary is present for both metrics and scenarios. It is still unclear whether and in which cases the intervention worked and you would be able to check if you agree with it or not.
Anonymous user, Anonymous user, Anonymous user

Clear - The summary is present for both metrics and scenarios. Overall it is still unclear whether and in which cases the intervention worked.
Anonymous user, Anonymous user, Anonymous user

Missing or confused - The summary is missing or confused. You would not be able to explain even the high level finding of the presented results.
Anonymous user

4 comments



Anonymous user
24-01-2025 23:32:58

does not exist

1 reply



O.J.B. Timmers (Ole)
29-01-2025 17:29:36

Not in the results section, will put it there as well, thanks.



Anonymous user
24-01-2025 05:38:48

While mostly being accurate, there is slight inconsistency present, regarding the summary for the impact of different scenarios (no insights)

1 reply



O.J.B. Timmers (Ole)
29-01-2025 17:29:57

Good point, will take that into account.



Anonymous user
24-01-2025 19:20:29

@page 6

This is a good conclusion based on the table data

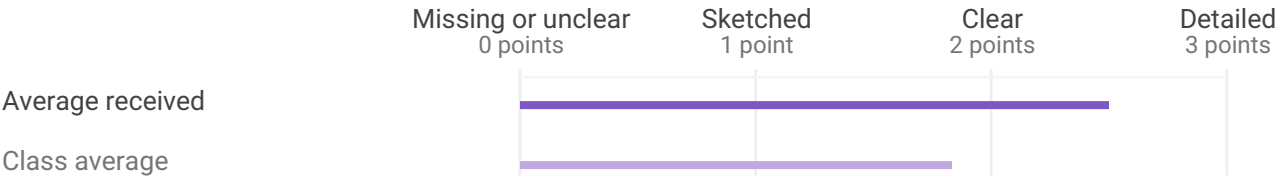


Anonymous user
24-01-2025 01:18:05

Concise summary

Final Summary

Rubric criterion



Received ratings

- Detailed - The final findings are clearly spelled out. It is clear on how they relate to what has been discussed in both the demographics and the results sections. You would be able to pitch the results in the elevator.
Anonymous user, Anonymous user

Clear - The final findings are clearly spelled out. There might still be some unclarity on how they relate to what has been discussed in the demographics or the results sections.
Anonymous user, Anonymous user

4 comments



Anonymous user
24-01-2025 23:34:12

While the final summary restates the intervention’s practical significance, it does not address its statistical limitations. This omission weakens the overall conclusions.



Anonymous user
24-01-2025 05:39:32

The summary is detailed and manages to capture all the points of the experiment



Anonymous user
24-01-2025 19:21:56

@page 6

What does practically significant mean?

1 reply



O.J.B. Timmers (Ole)
29-01-2025 17:30:47

Over $-+5\%$, commonly used to say something is significantly.

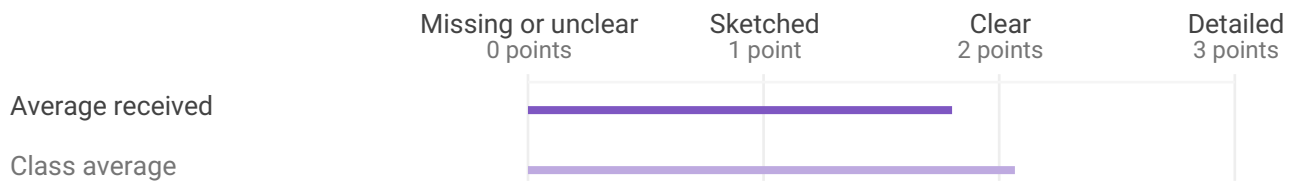


Anonymous user
24-01-2025 01:18:42

Great summary

Artefact

Rubric criterion



Received ratings

Detailed - There is a list of objectes and it is clear how they relate to what haves been computed. You would not need to actually look at the Github link.
Anonymous user

Clear - There is a list of objectes and for most of them it is clear how they relate to what haves been computed.
Anonymous user

Sketched - There is a list of objectes but you would not exactly know how they related to what haves been computed.
Anonymous user, Anonymous user

4 comments



Anonymous user
24-01-2025 23:36:06

The artifacts are well-documented but it would be insightful if the links are shared

1 reply



O.J.B. Timmers (Ole)
29-01-2025 17:31:27

That are put in now.



Anonymous user
24-01-2025 05:40:06

All artefacts are listed and their purpose is clear



Anonymous user
24-01-2025 19:22:58

Maybe a bullet list of objects and descriptions would be better than explaining where they are stored.

1 reply



O.J.B. Timmers (Ole)
29-01-2025 17:31:40

Will think about it.



Anonymous user
24-01-2025 01:18:26

Exact objects aren't listed