

Peer Review

I cannot review this paper until the data is statistically analyzed. For this, I will need to see actual scores, ranges, standard deviations, p-values, for all the attributes tested. Once this manuscript is revised and presents statistical data and analysis, it can be reviewed.

Revised manuscript explained

I have sent the revised version of my manuscript, and below is an explanation of what I did by section.

Participants

I moved the demographics on participants from results to the section on participants, adding the mean age of all participants and the standard deviation for all participants. As there is data per grade on gender, best language spoken, and reading habits, additional statistical data on each of these specific topics may become excessive; however, please let me know if you nonetheless wish for further statistical data in these areas.

Results

I moved the part explaining how the results were obtained from procedure to results, and so it is now located in the first paragraph of the results section, and it is more precise. In the second paragraph, I added the standard deviation and the range of my overall result/total mean score (the 2.33). In the third paragraph, for each component of empathetic engagement, where I had only provided the mean before, I added the standard deviation and range.

Discussion

In the second paragraph, I added the p-value, which was < 0.0001 , rejecting the null hypothesis.

I have therefore added ranges, standard deviations, and a p-value based on your comments. When you mention actual scores, the scores essential in understanding the analysis/discussion section and conclusions are provided— these are the empathetic engagement mean per grade, mean per engagement component, etc. I am nonetheless glad to add any additional scores asked for, or any extra statistical data if what I have added is insufficient. While editing, for increased clarity, I also added a sentence in the end of the implications section, added the literature review as a header in bold (after the introduction), and slightly edited some sentences.

Thank you for addressing my comments. however, I need to see individual scores for all the participants either in the body of the text or as a supplementary file. Please provide raw data.

Thank you for addressing the reviewer's comments. I have several comments on your revised manuscript.

1. I am not sure I agree with your premise that scores above average indicate hypothesis confirmation whereas those below do not (reference 14 notwithstanding). This average is arbitrarily chosen. What you need is a control group. I realize it is too late for this. Hence, I took your data and performed a multivariable linear regression on it by using the three variables of NU, NP and LD as independent and adding the CE and AE and using that as dependent. Per the results, the variable that most influences the sum of these empathies is NU, followed by NP which influences it negatively (although that may refute your hypothesis). LD does not have much influence. See tab2 of the attached spreadsheet. Therefore, I suggest you use 9 and 12th grades as controls (combine the two) and use the results from the 10th and 11th grades separately as test data. You will hence have mean, SD and # of students for control and mean, SD and # of students for test (for two populations, 10th and 11th). This will give you claimable data to show if there is or is not a difference in combined empathy based on NU and NP. I realize that this commits the inescapable fallacy of not having predetermined statistical parameters; but you can justify this as being an adaptive experiment were a difference between NU and NP would allow for post-experiment segregation/classification into control and test groups.

2. There may be other ways of obtaining a control group and/or categorizing the data. If you can think of any, please recalculate, reanalyze and present the justification, discussion and explanation in the manuscript.

3. My average in sheet 1 of 2.73 does not agree with yours (2.61). Please double check all your numbers and calculations.

I tried to categorize the data based on the reviewer's suggestion. In the results section, I explain that I divided the data into three groups post-hoc and I give the calculated data on mean and standard deviation for NU, NP, LD, AE, and CE for each group. I then give the results of the multivariable linear regression that would find which variables were most influential on combined empathy (NU and NP).

In the discussion, I interpret the numbers provided from the results section to refute my hypothesis and then validate the other two, based on what was found. Because of the changes done, I also removed figure 2 as it was no longer pertinent in the discussion, and I replaced it with a bar chart representing the control group's mean scores. In the limitations, I acknowledge the lack of predetermined statistical parameters and justify it as the reviewer suggested to. Lastly I changed a part of the conclusion regarding the results of the experiment.

I am not getting the same results you are. See attached spreadsheet. My averages, SD etc. for all the grades are different, hence is all the analysis. Please double check your numbers against mine.

I double checked my numbers, which were faulty. I replaced all of my results with those of the reviewer's. In Results: I provide the reviewer's spreadsheet's means and SDs for the empathetic engagement components (NU, NP, and AE+CE) for each grade and include the bar chart. Then I give the reviewer's results of 9+12 and 10+11 grades. To be able to validate or refute my main hypothesis, I calculated the comparison between 9+12 control and 10+11 test group to see if the difference is statistically significant (I hope the process and results for that are correct). Afterwards, I give the results of the multivariable linear regression, in which I hope my interpretation of the spreadsheet is correct. In Discussion: Based on the new results, I refuted my main hypothesis, as well as hypothesis A (based on the spreadsheet's means for NU and AE+CE per grade, there was no correlation, even though there appears to be a more negative influence as low NU creates high AE+CE). I also refute hypothesis B as the spreadsheet's p-value of 0.08 is greater than 0.05 making it statistically insignificant. I also had to alter some of the discussion on demographics. Lastly, I edited the conclusion to say that the main hypothesis, hypothesis A, and hypothesis B are refuted, giving the new values. I hope this was all done correctly. Thank you.

Thank you for addressing my comments. I have some more; hopefully the last of them.

1. Much of the data is better presented as tables. For example, the entire text under the heading "participants", all the bullet pointed results.
 2. Also, please put back in your previous analysis using the half-point (of 2.5?) as the cut off between control and test and calculate if there is any statistical difference using this method. You will then have two methods with which you will have compared the results.
 3. Can you add more references, depth to the explanation of the results? For example, why is there an inverse correlation between NU and AE+CE? Using anecdotal data, it does seem that the more one understands about a person; the less empathy one feels about that person... in loose correlation with the adage "familiarity breeds contempt". What does the published literature say about easy to understand modernist novels or genre versus the level of empathy they evoke? Do easier to understand fictional pieces evoke less empathy? Conversely, there is a direct correlation between NP and AE+CE, which means that the more one is immersed in the novel, the more the empathy. Also note that NP is inversely correlated to NU - which is at odds with the preceding explanation. Since, NP(greater immersion)=greater empathy, and greater immersion is only possible with greater understanding (NU) hence NU must be directly proportional to NP; but it is not. The purpose of this meandering review is to open a dialog into conflicting results.
-

In my hypotheses, for hypothesis A, I specify that there will be a direct correlation, not just any correlation (as the foundation for my hypothesis is Penner, Erin Kay et al's pattern which claims increase in NU begets increase in engagement) as I had not done that before. Under participants, I put all of the participants' demographics into a table (table 1). Under results, all the bullet pointed results are in tables. Table 2 shows the means and standard deviations per grade for NU, NP, and AE + CE. Table 3 shows the results for the control group and test group. I also added a 4th table to represent the multivariable linear regression's results for NU and NP. In Discussion: After the first comparison, I added the comparison with the half-point as a cut off and added the contingency table used— how many participants were above or below 2.5. I hope the calculation for this is correct. Ultimately, both methods refute the hypothesis. Afterwards, I elaborate

on NU and AE +CE based on what I found in the existing literature— the effect of high NU remains undiscussed, there's mainly just the idea that there needs to be a minimum of NU. I add in Future Direction that this is a possible point of future study. I explain the inverse correlation between NU and NP with the note on modernist unrealism detracting from NP, as more NU means greater understanding of the unrealism, leading to lower NP. To explain more, I also add a point on reference 11's significance of spatiotemporal framework and connect it back to the idea of aesthetically distanced reading.

Accepted