

PEER REVIEW HISTORY

BMJ Open publishes all reviews undertaken for accepted manuscripts. Reviewers are asked to complete a checklist review form (<http://bmjopen.bmj.com/site/about/resources/checklist.pdf>) and are provided with free text boxes to elaborate on their assessment. These free text comments are reproduced below.

ARTICLE DETAILS

TITLE (PROVISIONAL)	The effect of a 2-day communication skills training on nursing and midwifery students' empathy: a randomised controlled trial
AUTHORS	Alhassan, Mustapha

VERSION 1 – REVIEW

REVIEWER	Pilar Bas-Sarmiento University of Cadiz, Spain
REVIEW RETURNED	05-Jun-2018

GENERAL COMMENTS	<p>Congratulations for the study. As a strength it is a randomized controlled study conducted in Ghana. The article is presented in a clear way, is based on a theoretical model and the discussion provides updated bibliography.</p> <p>Abstract</p> <p>When summarizing the results the author says: “there was no statistically significant difference in the scores of empathy between the groups. Eventhough there were slight increases in the intervention group from baseline”. If we see the scores that were given on the baseline and the scores between the groups (between the posttest and follow-up groups), this statement does not seem very convenient.</p> <p>Metodology</p> <p>One of the limitations of this type of studies, showed in systematic reviews, is that training is not detailed, which prevents it from being replicated, as in this case. Therefore, the intervention should be specified. In fact, the use of role-playing methodology is reported in the strengths. However, this learning tool is not specified in the description of the communication skills training.</p> <p>At the methodological level it is necessary to clarify if the same people who carried out the training were who analyzed the data and if they were blinded for the analysis.</p> <p>In the randomisation section the author says: “The NMS were separated before random assignment to ensure that both professions were aproximately equally represented in the groups”. However, when the characteristics of the sample are defined, a non equitable distribution appears: 31 students for intervention group and 11 for control group, in midwifery students.</p> <p>Although the analysis of variance is robust for slight deviations from normality, the analysis of data does not specify if the distribution of the variable was analyzed to verify normality.</p> <p>Outcomes</p> <p>As recommended by the Consort guide, a flow chart should be made</p>
-------------------------	---

	<p>specifying the participants and the date of recruitment, the participants who were assigned to each group, those who finally received the intervention and continued during the follow-up and those who were finally analyzed (as the data show, it seems that there were participants that were lost).</p> <p>Discussion and conclusions</p> <p>In the discussion it would be necessary to distinguish, with regard to the studies that are specified (those that have been successful and those that did not achieve a change in empathy) the circumstances regarding to training, measures to be evaluated, forms of evaluation, etc.</p> <p>The conclusion needs to be reformulated. It could be that the training is not appropriate or the training time is not enough (previous studies have shown that excessively short trainings do not succeed) or the measurement used to evaluate does not reflect the changes. In fact the Jefferson is a general measure of self-perceived empathy that remains fairly stable. You could differentiate between wanting to empathize and empathize, that is, the training can improve the ability to empathize but not the motivation to empathize or vice versa.</p> <p>Likewise, we do not know if they had the opportunity to demonstrate what they learned in a simulated test or in reality. It would have been wise to include another type of measure as external observers and / or the perception of the patient and to analyze if they displayed this competences. In fact, in previous studies you can see how all the measures of this kind change except the Jefferson (Bas-Sarmiento, P. et al., 2017. Efficacy of empathy training in nursing students: A quasi-experimental study. Nurse Education Today 59. DOI: 10.1016/j.nedt.2017.08.012; Bas P. et al., Teaching Empathy to Nursing Students: A Randomised Controlled Trial. International Journal of Nursing Studies. In review).</p>
--	--

REVIEWER	Evonne Kaplan-Liss MD MPH TCU & UNTHSC School of Medicine USA
REVIEW RETURNED	09-Jun-2018

GENERAL COMMENTS	<p>1. What CST or intervention did the control group receive ? not clear</p> <p>2."Participants were informed of the objectives of the study and were also given opportunity to ask questions for a better understanding of the study. " The participants were made aware of empathy being a outcome of this study and since JES is self-reported, it may have impacted their self-report. Selection bias may have impacted the lack of significance. It's possible that participants that volunteered were more empathetic compared to baseline and JES is self-report.</p> <p>3. The CST training explanation should be more detailed so it could be replicated and compared and contrasted to other studies mentioned in the discussion. The discussion mentions that their results differ than many published studies but not all empathy or CST are alike - and the mode of delivery is also important to differentiate</p> <p>5. The control and intervention groups demographics were dissimilar (table 2) even though randomized. "The NMS were separated before random assignment to ensure that both professions were approximately equally represented in the groups. " but they weren't (33% v 11% midwifery students in intervention v. control respectively)</p>
-------------------------	---

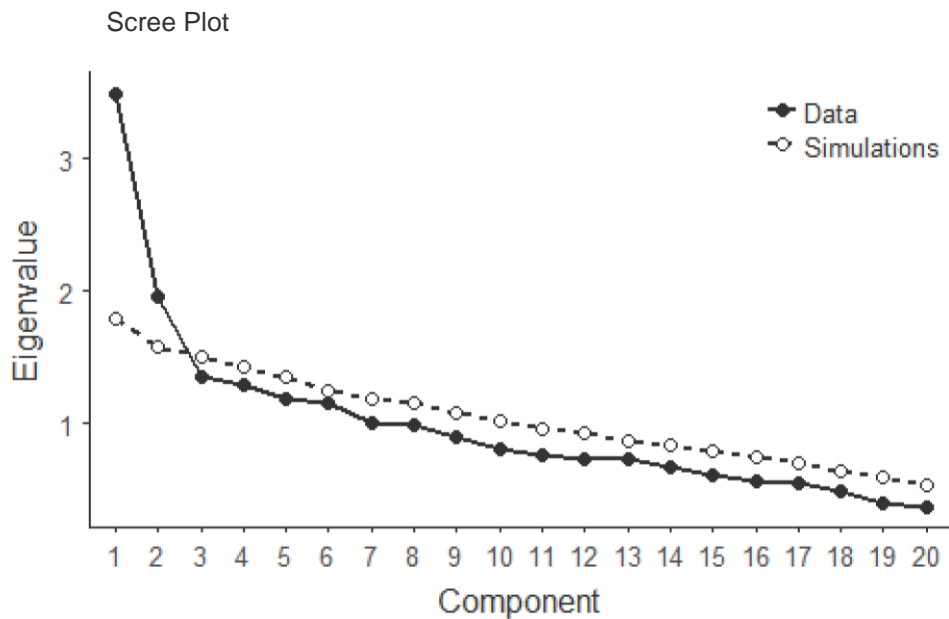
REVIEWER	James Green
-----------------	-------------

	University of Limerick Ireland
REVIEW RETURNED	08-Aug-2018

GENERAL COMMENTS	<p>This seems a worthwhile if ambitious attempt to measure changes in empathy following two days of communication skills training.</p> <p>Thank you for providing the raw data on request. I must apologise for the delay in providing this review. Due to some email server configurations issues, I didn't receive the data for several weeks.</p> <p>First up, as with many other educational interventions, it seems ambitious to expect to see meaningful change in what might expected to be a relatively stable trait after only a short amount of training. However, reporting null findings is useful to illustrate that this is the case.</p> <p>Having reviewed the paper, and also looked through the data, I have two main points.</p> <ol style="list-style-type: none"> 1. As currently described, the analyses that you actually reported don't bear much resemblance to what I thought that you had done based on my reading of the paper. 2. Scales are no longer assumed to have inherent validity – that is, validity should be assessed with each new study, rather than relying on the original analysis of validity. This is particularly relevant where a scale has been translated which I presume is the case here. Using the data, I was thus able to perform an exploratory principal components analysis (see below), and found clear evidence of two factors, that seem to map on to two of the three factors usually reported with this scale. I therefore recommend that you consider re-analysing your data considering the multi-dimensional nature of the scale. For some recent readings on why this is important, and what else you might consider doing, the following may be useful: <p>Crutzen and Peters, 2015 http://dx.doi.org/10.1080/17437199.2015.1124240 Dima 2018 https://www.tandfonline.com/doi/pdf/10.1080/21642850.2018.1472602 As I will outline later, two new free and easy-to-use software packages, JASP and JAMOVI may offer you more powerful and modern analyses than SPSS. At least one of those also allows for confirmatory factor analysis.</p> <p>Principal Component Analysis</p> <p style="padding-left: 40px;">Component Loadings</p> <hr/> <table style="margin-left: auto; margin-right: auto; border-collapse: collapse;"> <thead> <tr> <th colspan="3" style="text-align: center; border-bottom: 1px solid black;">Component</th> </tr> <tr> <th style="text-align: center; border-bottom: 1px solid black;">1</th> <th style="text-align: center; border-bottom: 1px solid black;">2</th> <th style="text-align: center; border-bottom: 1px solid black;">Uniqueness</th> </tr> </thead> <tbody> <tr> <td style="border-top: 1px solid black; border-bottom: 1px solid black;"> </td> <td style="border-top: 1px solid black; border-bottom: 1px solid black;"> </td> <td style="border-top: 1px solid black; border-bottom: 1px solid black;"> </td> </tr> </tbody> </table>	Component			1	2	Uniqueness			
Component										
1	2	Uniqueness								

JSE_T1_Qn1		0.553	0.701
JSE_T1_Qn2	0.446		0.798
JSE_T1_Qn3		0.360	0.848
JSE_T1_Qn4	0.551		0.596
Component Loadings			
Component			
	1	2	Uniqueness
JSE_T1_Qn5	0.535		0.713
JSE_T1_Qn6		0.432	0.766
JSE_T1_Qn7		0.360	0.825
JSE_T1_Qn8		0.499	0.756
JSE_T1_Qn9	0.342	0.351	0.725
JSE_T1_Qn10	0.485		0.735
JSE_T1_Qn11		0.552	0.638
JSE_T1_Qn12		0.622	0.605
JSE_T1_Qn13	0.477		0.767
JSE_T1_Qn14		0.379	0.842
JSE_T1_Qn15	0.591		0.657
JSE_T1_Qn16	0.676		0.526
JSE_T1_Qn17	0.321		0.886
JSE_T1_Qn18	-0.551		0.703
JSE_T1_Qn19			0.927
JSE_T1_Qn20	0.689		0.535
Note. 'oblimin' rotation was used			

Eigenvalues



Other points

1. Consider publicly storing the data on a scientific data repository site
2. I can't replicate the power analysis. No power value is given (80%, 90%, 95%?), but using $d = 0.25$ and $\alpha = .05$, and a variety of normal power values, I did not get $n = 197$. Please describe your power analysis in further detail (power level, what type of test, what software)
3. The description of the analysis is unclear. Before I had received the data and your output files, I had been speculating that you had in fact run simple two group ANOVA (equivalent to an independent t-test) at each of the three time points. So this needs to be much carefully described as a 2 (intervention condition, between) x 3 (time, repeated) design. However, a more common way to analyse and intervention would be ANCOVA with Time 2 as dependent variable, Time 1 as covariate, and intervention condition as a between factor. However, given the lack of effect/difference, no analytic change is going to change the essence of your results.
4. Alternately, given that there were two health professional student groups (nurses and midwives), it might be possible to run a more complex design with a factor representing profession, and also potentially nesting students within session, in case there are differences in teaching between sessions, as well as controlling for baseline. I briefly explored this, but it didn't appear there was any difference.
5. No effect sizes are reported, either for within group change, or for the main comparison of interest between conditions at each time point. Either Cohen's D or preferable Hedges g should be reported (the latter is unbiased), with confidence intervals. Jamovi or JASP, two new free alternative to SPSS calculated both, or Excel spreadsheets exist to make these calculations. These effect sizes could be added to Table 3. Or you could look at effect sizes for the change over time.

Independent Samples T-Test						
	t	df	p	Cohen's d	95% CI for Cohen's d	
					Lower	Upper
Time 1	1.051	172.0	0.295	0.160	-0.139	0.458
Time 2	1.323	172.0	0.187	0.201	-0.098	0.499
Time 3	-1.658	172.0	0.099	-0.252	-0.551	0.048

Note. Student's t-test.

6. What language was the questionnaire administered in? If it was translated, how, and what pre-testing/quality control was conducted to ensure accuracy of the translation.
7. Table 2 – given the margin of error for a sample of 173 is ~5%, rounding to no (zero) decimal places is appropriate. Table 3 – one decimal place is probably appropriate.
8. There appear to be some quite extreme differences between intervention and control groups (eg gender, religion). Comment on as a limitation?
9. 'Do you have children' is redundant with the 'Number of Children' section in the table
10. Blinding — were all tutors absolutely unaware of the design, given that the second sessions of training were obviously the control group, if people were aware of the design.
11. Demand effects. At line 112, you told participants the objectives of the study. This may have changed participants behaviour (social desirability/demand effects, also sometimes inaccurately referred to as the Hawthorne effect). Clarify the extent to which participants were explicitly informed of the objectives, and perhaps next time consider being much less specific, providing fuller explanation after data collection has concluded.
12. As per the arguments of Stephen Pinker, cognitive psycholinguist, abbreviations provide minor savings to the author, but steal time and mental effort from the reader. I would recommend spelling out CST as either training or communication skills training throughout the manuscript, and similarly for other abbreviations (eg NMS).
13. The text at lines 2013-215 seems largely redundant with the table, unless you were to put effect sizes for the changes here

VERSION 1 – AUTHOR RESPONSE

Reviewer: 1
 Reviewer Name: Pilar Bas-Sarmiento
 Institution and Country: University of Cadiz, Spain

Reviewer 1
 Please state any competing interests or state 'None declared': None declared

Author revision
 Page 12, Line 452: "None declared" has been stated under competing interest.

Reviewer 1

Abstract

When summarizing the results the author says: “there was no statistically significant difference in the scores of empathy between the groups. Eventhough there were slight increases in the intervention group from baseline”. If we see the scores that were given on the baseline and the scores between the groups (between the posttest and follow-up groups), this statement does not seem very convenient.

Author revision

Page 1, Line 19 - 21: It has been revised to read as “The intervention group had baseline - T1 (M = 109.75; SD = 9.76), and post-test - T2 (M = 111.85; SD = 8.95), whereas the control group had baseline - T1 (M = 107.93; SD = 11.46); and post-test - T2 (M = 110.01; SD = 11.03)”.

Reviewer 1

Methodology

One of the limitations of this type of studies, showed in systematic reviews, is that training is not detailed, which prevents it from being replicated, as in this case. Therefore, the intervention should be specified. In fact, the use of role-playing methodology is reported in the strengths. However, this learning tool is not specified in the description of the communication skills training.

Author revision

Page 4, Line 174: The Communication Skills Training (CST actually included role-playing. “Role Playing” has been included in the article at Page 5, Line 192.

Reviewer 1

At the methodological level it is necessary to clarify if the same people who carried out the training were who analyzed the data and if they were blinded for the analysis.

Author revision

Page 4, Line 172: It was only the author (MA) who did the data analysis. There was no blinding during the data analysis. The following sentence has been inserted to take care of that “The data was analysed by MA without blinding”.

Reviewer: 1

In the randomisation section the author says: “The NMS were separated before random assignment to ensure that both professions were approximately equally represented in the groups”. However, when the characteristics of the sample are defined, a non equitable distribution appears: 31 students for intervention group and 11 for control group, in midwifery students.

Author revision

Page 5, Table 2: The nursing student were 131 (Intervention Group 62 and Control Group 69), while the Midwifery students were 42 (Intervention 31 and Control 11). Due to this uneven ratio of Nursing Students (75.72%) to Midwifery (24.28%), the separation was to ensure both groups had approximate ratio in the Groups.

Reviewer 1

Although the analysis of variance is robust for slight deviations from normality, the analysis of data does not specify if the distribution of the variable was analyzed to verify normality.

Author revision

Page 4 Line 157 – 159: The distribution of variable was analysed for normality and I have include it in the Article as “A Shapiro-Wilk’s test ($p < .05$) [16,17] and a visual inspection of their histograms

showed variable scores were approximately normally distributed”.

Reviewer 1

Outcomes

As recommended by the Consort guide, a flow chart should be made specifying the participants and the date of recruitment, the participants who were assigned to each group, those who finally received the intervention and continued during the follow-up and those who were finally analyzed (as the data show, it seems that there were participants that were lost).

Author revision

In Supplementary Material Fig. 1: Flowchart:

The dates for recruitment (15 June 2013), Baseline test for Intervention and Control Groups (15 June 2013), Post-test for Intervention Group (17 June 2013), Post-test for Control Group (18 June 2013), and Follow-up test for both Intervention and Control Groups (12 December 2013) have been included in the Flowchart.

The lost in participants was as a result of incomplete data and outcome measures not returned. The number of incomplete and outcome measures not returned are indicated in Fig. 1: Flowchart.

Reviewer 1

Discussion and conclusions

In the discussion it would be necessary to distinguish, with regard to the studies that are specified (those that have been successful and those that did not achieve a change in empathy) the circumstances regarding to training, measures to be evaluated, forms of evaluation, etc.

Discussion and conclusions In the discussion it would be necessary to distinguish, with regard to the studies that are specified (those that have been successful and those that did not achieve a change in empathy) the circumstances regarding to training, measures to be evaluated, forms of evaluation, etc.

Author revision

Page 7, Line 223 – 234 and Line 239 – 240: I have indicated in the discussion 4 studies that showed enhancement of empathy in Page 7, Line 248 – 249 “The findings from this study are in contrast to the findings from a similar study that showed enhancement of empathy in nurses [20–23]”.

Nine other studies thought the effectiveness of empathy training programmes “Research has shown that there are a number of studies that doubt the effectiveness of empathy training programmes in nursing education and rather reported stability in empathy[24–32]”.

Reviewer 1

The conclusion needs to be reformulated. It could be that the training is not appropriate or the training time is not enough (previous studies have shown that excessively short trainings do not succeed) or the measurement used to evaluate does not reflect the changes. In fact the Jefferson is a general measure of self-perceived empathy that remains fairly stable. You could differentiate between wanting to empathize and empathize, that is, the training can improve the ability to empathize but not the motivation to empathize or vice versa. Likewise, we do not know if they had the opportunity to demonstrate what they learned in a simulated test or in reality. It would have been wise to include another type of measure as external observers and / or the perception of the patient and to analyze if they displayed this competences. In fact, in previous studies you can see how all the measures of this kind change except the Jefferson (Bas-Sarmiento, P. et al., 2017. Efficacy of empathy training in nursing students: A quasi-experimental study. *Nurse Education Today* 59. DOI: 10.1016/j.nedt.2017.08.012; Bas P. et al., Teaching Empathy to Nursing Students: A Randomised Controlled Trial. *International Journal of Nursing Studies*. In review).

Author revision

Page 8, Line 298 – 303: I agree perfectly that the conclusion needs to be reformulated. I have therefore included 2 sentences

(1) "The participants were made aware of empathy being an outcome of this study and since JES is self-reported, it may have impacted their self-report. Selection bias may have impacted the lack of significance. It's possible that participants that volunteered were more empathetic compared to baseline and JES is self-report".

(2) "More so, the 2-day training time was not enough and that could have accounted for none enhancement of empathy".

Thank you very much for referring me to the 2 previous studies.

I wish to thank you for the use of your valuable time to review and make important suggestions.

Reviewer: 2

Reviewer Name: Evonne Kaplan-Liss MD MPH Institution and Country: TCU & UNTHSC School of Medicine, USA

Reviewer 2

Please state any competing interests or state 'None declared': None declared.

Author revision

Page 12, Line 452: "None declared" has been stated under competing interest.

Reviewer 2

1. What CST or intervention did the control group receive? not clear

Author Revision

Page 3, Line 119: The intervention and control group both had the same CST but at different time points. The sentence "The CST for both groups were the same" has been inserted to take care of that.

Reviewer 2

2. "Participants were informed of the objectives of the study and were also given opportunity to ask questions for a better understanding of the study". The participants were made aware of empathy being a outcome of this study and since JES is self-reported, it may have impacted their self-report. Selection bias may have impacted the lack of significance. It's possible that participants that volunteered were more empathetic compared to baseline and JES is self-report.

Author Revision

Page 8, Line 298 – 303: I agree perfectly with you that selection bias may have impacted the lack of significance. I have therefore included your suggestion as a limitation of the study. Thank you very much. I have inserted two sentences:

(1) "The participants were made aware of empathy being an outcome of this study and since JES is

self-reported, it may have impacted their self-report. Selection bias may have impacted the lack of significance. It's possible that participants that volunteered were more empathetic compared to baseline and JES is self-report".

(2) "More so, the 2-day training time was not enough and that could have accounted for none enhancement of empathy".

3. The CST training explanation should be more detailed so it could be replicated and compared and contrasted to other studies mentioned in the discussion. The discussion mentions that their results differ than many published studies but not all empathy or CST are alike - and the mode of delivery is also important to differentiate

Author Revision

The CST training that was developed was comprehensive and I am ready to share it with anyone who is interested in replicating the study. Contact Email mustaph@uds.edu.gh

5. The control and intervention groups demographics were dissimilar (Table 2) even though randomized. "The NMS were separated before random assignment to ensure that both professions were approximately equally represented in the groups." but they weren't (33% V 11% midwifery students in intervention v. control respectively)

Author revision

Page 5, Table 2: The nursing student were 131 (Intervention Group 62 and Control Group 69), while the Midwifery students were 42 (Intervention 31 and Control 11). Due to this uneven ratio of Nursing students (75.72%) to Midwifery students (24.28%), the separation was to ensure both groups had approximate ratio in the Groups.

I wish to thank you for the use of your valuable time to review and make important suggestions.

Reviewer: 3

Reviewer Name: James Green

Institution and Country: University of Limerick, Ireland

Reviewer 3

Please state any competing interests or state 'None declared': None declared

Author revision

Page 12, Line 452: "None declared" has been stated under Competing Interest.

VERSION 2 – REVIEW

REVIEWER	James Green University of Limerick, Ireland
REVIEW RETURNED	24-Sep-2018

GENERAL COMMENTS	Thank you for the careful revision of the manuscript, and I am sorry that my previous comments as an attachment were not very obvious. I am particularly pleased that you have made the data available on dryad, and I hope that you find the new software packages useful. I still have a couple of comments. 1. I used G-Power to check your prior power analyses, and with
-------------------------	---

	<p>believe that you have inadvertently selected an inappropriate model “t-test::Correlation: point biserial”, and entered a rho value of 0.25, which yields your sample size of $n = 197$. However, using a more appropriate model (either t-test::difference between independent groups, or ANOVA::between/repeated) suggests that you are still adequately powered for at least a medium effect.</p> <p>2. The additional analyses in the supplementary materials are not mentioned in the main paper, and should be.</p> <p>3. My original goal for the PCA is that ideally you should treating each component as a dependent variable, however on reinspection just now, I’ve just realised that the second component is all the reverse-scored items. This suggests that perhaps the scale is not working as intended, and may undermine its validity. However, it seems a common enough problem, so may be OK.</p> <p>4. Related to that, if the JSE was administered in English, you should mention the English fluency of the students, which language they are taught in, and which language they are most likely to practice in.</p> <p>5. The means for intervention and control are presented in the abstract; I would present Cohen’s d there as well.</p> <p>More minor</p> <p>1. You use the greek letter rho (ρ) when you mean to use an italicised p. They look similar, but rho has other specific meanings</p> <p>2. I’d reiterate my prior point on rounding for Table 3. The differences observed were very small, and having the extra decimal places makes these values seem far more precise than they really are (ie measurement error is high and the mean differences are low, so the small apparent differences are not very meaningful).</p>
--	--

VERSION 2 – AUTHOR RESPONSE

Reviewer: 3

Reviewer Name: James Green

Institution and Country: University of Limerick, Ireland

Reviewer’s comments to Author

Please state any competing interests or state ‘None declared’: None declared.

Author revision and response

Page 14, Line 508: under competing interest, I have stated “None declared”

Reviewer’s comments to Author

Please leave your comments for the authors below

Thank you for the careful revision of the manuscript, and I am sorry that my previous comments as an attachment were not very obvious. I am particularly pleased that you have made the data available on dryad, and I hope that you find the new software packages useful.

Author revision and response

Thank you very much for the new software you suggested. I have revised the manuscript to reflect the comments and suggestions you made. I have discussed with colleagues how the new software you suggested broadened my understating and how more powerful and modern these softwares are in analysing data a compared to SPSS.

Reviewer’s comments to Author

I still have a couple of comments.

1. I used G-Power to check your prior power analyses, and with believe that you have inadvertently selected an inappropriate model “t-test::Correlation: point biserial”, and entered a rho value of 0.25, which yields your sample size of $n = 197$. However, using a more appropriate model (either t-test::difference between independent groups, or ANOVA::between/repeated) suggests that you are still adequately powered for at least a medium effect.

Author revision and response

Great suggestion and thank you very much for checking the prior power analysis to determine the appropriateness. I am happy to note that, using a more appropriate model (either t-test::difference between independent groups, or ANOVA::between/repeated) suggested that I am still adequately powered for at least a medium effect.

Reviewer’s comments to Author

2. The additional analyses in the supplementary materials are not mentioned in the main paper, and should be.

Author revision and response

Page 5, Lines 187 – 189: The following sentence have been inserted “I have included the results as supplementary files in the form of principal component analysis, Chi-squared test, Independent t-test, and Scree plot (Additional file 1: Table 1, Table 2, Table 3 and Additional file 2: Fig. 1)”.

Page 13, Lines 490 – 494: the following have been included after the References

“Additional Files

Additional file 1: Table 1; Principal component analysis, Table 2; Chi-squared test, and Table 3; Independent samples t-test
Additional file 2: Fig. 1; Scree plot”

Reviewer’s comments to Author

3. My original goal for the PCA is that ideally you should treating each component as a dependent variable, however on reinspection just now, I’ve just realised that the second component is all the reverse-scored items. This suggests that perhaps the scale is not working as intended, and may undermine its validity. However, it seems a common enough problem, so may be OK.

Author revision and response

Well noted and thank you very much for observing this common problem.

Reviewer’s comments to Author

4. Related to that, if the JSE was administered in English, you should mention the English fluency of the students, which language they are taught in, and which language they are most likely to practice in.

Author revision and response

Page 4, Lines 153 -155: I have inserted this sentence in the manuscript “This is because the students are very fluent in English. They are taught in English language from primary school because English is the official language in Ghana, and they will practice in English.”

Reviewer’s comments to Author

5. The means for intervention and control are presented in the abstract; I would present Cohen’s d there as well.

Author revision and response

Page 1, Lines 34 – 37: I have included Cohen's d in the abstract

Reviewer's comments to Author

More minor

1. You use the greek letter rho (ρ) when you mean to use an italicised p. They look similar, but rho has other specific meanings.

Author revision and response

The Greek letter rho (ρ) has been changed to italicised p in the text in the following locations:

Page 1, Line 34,

Page 3, Line 101

Page 4, Line 179

Page 4, Line 185

Page 5, Line 191

Page 5, Line 192

Page 5, Line 194

Page 5, Line 196

Page 5, Line 197

Page 6, Line 237

Page 7, Line 243

Page 7, Line 257

Page 7, Line 266

Page 8, Line 284

Page 8, Line 289

Page 8, Line 297

Reviewer's comments to Author

2. I'd reiterate my prior point on rounding for Table 3. The differences observed were very small, and having the extra decimal places makes these values seem far more precise than they really are (ie measurement error is high and the mean differences are low, so the small apparent differences are not very meaningful).

Author revision and response

Page 6, Table 3, Lines 231 - 233: Please, I have changed the mean (M) and standard deviation (SD) values in Table 3 to one decimal place.

Based on the changes made in Table 3 to one decimal place, I have accordingly changed the figures in page 6, Lines 226 – 229 to match those changes.

Page 1, Lines 34 – 37: I have included Cohen's d in the abstract.

VERSION 3 – REVIEW

REVIEWER	James Green University of Limerick, Ireland
REVIEW RETURNED	27-Jan-2019
GENERAL COMMENTS	I am happy with the changes made in response to my last review.