

Interactive comment on ““Thanks for helping me find my enthusiasm for physics!” The lasting impacts “research in schools” projects can have on students, teachers, and schools” by Martin O. Archer and Jennifer DeWitt

Anonymous Referee #2

Received and published: 3 November 2020

This is an excellent paper, and describes important, and thorough research into the effects of an engagement programme - something which is notoriously hard to do. My comments are largely very minor

Section 3: In point 5, the authors mention reliability. It would be interesting to know how reliable they found the codes (i.e. what were different between the first and second coder, and was it significant)

4.1.1: (sentence 2) "Additionally they were asked to reassess their confidence before

C1

having undertaken the project." I found this ambiguous - reassessment implies a second assessment, when the text mentions that there was no pre-assessment. Perhaps "retrospectively assess" would be more accurate?

Figure 1: I very much like the figures in the article, and they are all clear. With this figure, it may be worth considering applying some transparency to the points as some are overlapping. I am, however, willing to believe that this makes it too confusing, but it is something the authors should consider.

Table 4: (caption) if 11 weren't placed in dimensions, it may be clearer to say using n=52 of 63 responses?

Figure 4: what intervals are the error bars (1-sigma, 95%?). It is worth noting that these intervals are not reliable with small n or low probabilities of success. I suggest at least an acknowledgement that these should be treated as indicative given the sample size.

Section 4.3: in the final paragraph there are details of (as yet) unpublished works which name the "anonymised" schools. Does this risk de-anonymising the schools used here, if the results are published in the future?

Section 5.2: The authors use a null hypothesis of 2. Would a better quantitative test to simply be to code positive vs negative, without the division into planned and definite? (i.e. give definite a score of 2 as well for this purpose, with a null hypothesis of 1.5)? At present the null hypothesis is that 2/3 of respondents claim a positive impact.

Figure 5: I found the grey error bars hard to spot, as they are narrow and overlap the black error bars. Perhaps thicker lines and/or offset horizontally with respect to the black error bars?

Interactive comment on Geosci. Commun. Discuss., https://doi.org/10.5194/gc-2020-36, 2020.

C2